Deformable model fitting by regularized landmark mean-shift

Daniel Šaranović

MACHINE LEARNING AND APPLICATIONS GROUP

April, 2018

Pregled predavanja I

Pregled predavanja II

Eksperimenti

Pitanja

Pregled predavanja

Zasto 'Deformable'

- Jedan vid traženja objekata u slikama jestu Template based algoritmi.

Zasto 'Deformable'

- Jedan vid traženja objekata u slikama jestu Template based algoritmi.
- Zgodno ako su objekti rigidni - tj. ne menjaju oblik
- Algoritmi bazirani na 'Savitljivom' modelu traže specifične tačke na objektima koji uglavnom nisu rigidni.

Deformable modeli u akciji

- Deformable modeli mogu da se koriste za segmentaciju lica i nalaženje ključnih tačaka lica.

Deformable modeli u akciji

- Deformable modeli mogu da se koriste za segmentaciju lica i nalaženje ključnih tačaka lica.
- Ako bi hteli da automatizujemo tumačenje gestova ili govor tela potrebno nam je fino praćenje delova tela i lica.

Deformable modeli u akciji

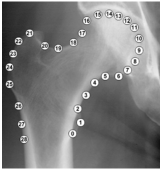
- Deformable modeli mogu da se koriste za segmentaciju lica i nalaženje ključnih tačaka lica.
- Ako bi hteli da automatizujemo tumačenje gestova ili govor tela potrebno nam je fino praćenje delova tela i lica.
- Segmentacija medicinskih slika.

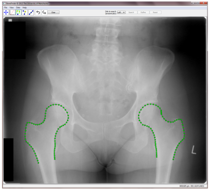
Deformable modeli u akciji

- Deformable modeli mogu da se koriste za segmentaciju lica i nalaženje ključnih tačaka lica.
- Ako bi hteli da automatizujemo tumačenje gestova ili govor tela potrebno nam je fino praćenje delova tela i lica.
- Segmentacija medicinskih slika.
- Automatska segmentacija zuba.

Još par primera

Osnovna ideja jeste da kreiramo trening skup sa označenim ključnim tačkama koji zatim koristimo da istreniramo naš 'Savitljivi model' pomoću kojeg nalazimo te iste tačke na slikama od interesa.





Primer iz medicine

- Sve je to lepo , je l' leči i rak?

Primer iz medicine

- Sve je to lepo , je l' leči i rak?
- Pa, na neki način i da.

Pregled predavanja

ASM

Istorijski pregled

ASM Predprocesiranje

PCA

Active Shape Model

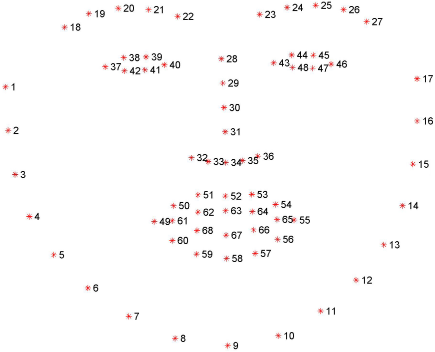
ASM
Uvod

- Active Shape Model (Cootes, Taylor 1992) je jedan od prvih značajnijih radova na temu Deformable modela.
- Drugi naziv je 'Smart snakes'
- Ideja je da imamo veliku količinu objekata sa uniformno označenim ključnim tačkama

ASM

Uvod

- Active Shape Model (Cootes, Taylor 1992) je jedan od prvih značajnijih radova na temu Deformable modela.
- Drugi naziv je 'Smart snakes'
- Ideja je da imamo veliku količinu objekata sa uniformno označenim ključnim tačkama
- Npr. ovako



PDM

- Point Distribution Model je uređen skup tačaka objekta.
- Tačke koje se nalaze u PDM-u nazivamo još i ključne tačke.
- Umesto da baratamo sa celim objektom radimo sa odgovarajućim PDM-om.

Pregled predavanja

ASM

Istorijski pregled

ASM Predprocesiranje

PCA

Active Shape Model

Data Manipulation

- Prvo želimo da izvršimo poravnanje naših instanci u trening skupu (Skaliranje, rotacija, translacija).

Data Manipulation

- Prvo želimo da izvršimo poravnanje naših instanci u trening skupu (Skaliranje, rotacija, translacija).
- Translacija: Od svakog PDM-a oduzmemo njegov prosek po x i y koordinati tako da su centrirani u $(0, 0)$

Data Manipulation

- Prvo želimo da izvršimo poravnanje naših instanci u trening skupu (Skaliranje, rotacija, translacija).
- Translacija: Od svakog PDM-a oduzmemo njegov prosek po x i y koordinati tako da su centrirani u $(0,0)$
- Uzimamo za reper prvi objekat (npr.) i ako je z_1 njegov PDM skaliramo tako da bude $\|z_1\| = 1$. (Samo podelimo sa njegovom normom sve elemente)

Data Manipulation

- Prvo želimo da izvršimo poravnanje naših instanci u trening skupu (Skaliranje, rotacija, translacija).
- Translacija: Od svakog PDM-a oduzmemo njegov prosek po x i y koordinati tako da su centrirani u $(0, 0)$
- Uzimamo za reper prvi objekat (npr.) i ako je z_1 njegov PDM skaliramo tako da bude $\|z_1\| = 1$. (Samo podelimo sa njegovom normom sve elemente)
- Skaliramo i rotiramo ostale oblike tako da se poravnaju sa z_1.
  Skaliramo oblik j za s_j i rotiramo za ugao θ_j.
- Ovo je još poznato i kao Procrustes analysis.

ASM

- Ako imamo n ključnih tačaka - svaki oblik nam leži u prostoru sa $2n$ dimenzija. Svakako da je moguće smanjiti dimenzionalnost (nisu sve koordinate nezavisne) - PCA

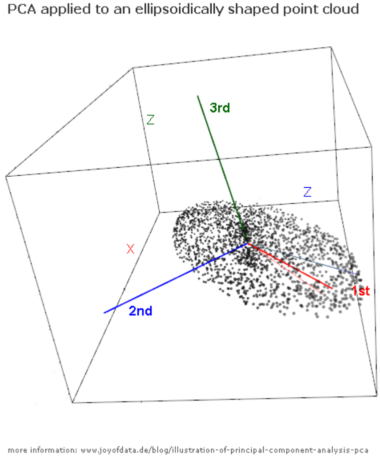
Pregled predavanja

PCA Podsetnik

- Nisu sve tačke nezavisne

PCA Podsetnik

- Nisu sve tačke nezavisne
- PCA nam omogućava da nađemo pravce po kome podaci najviše variraju.



Pregled predavanja

ASM

Istorijski pregled

ASM Predprocesiranje

PCA

Active Shape Model

ASM

- Dakle svaki objekat možemo da predstavimo kao

$$x = \mu + Pb$$

Gde je P ortogonalna matrica sopstvenih vektora dobijenih PCA analizom i b vektor koordinata u novom koordinatnom sistemu.

ASM

- Dakle svaki objekat možemo da predstavimo kao

$$x = \mu + Pb$$

  Gde je P ortogonalna matrica sopstvenih vektora dobijenih PCA analizom i b vektor koordinata u novom koordinatnom sistemu.

- Ako imamo date koordinate objekta u koordinatnom sistemu objekta (koordinate objekata kada je poravnat) x možemo da dobijemo koordinate objekta u koordinatnom sistemu slike X rotacijom, skaliranjem i transliranjem sa

$$X = M[s, \theta](x) + \boldsymbol{X}_c$$

  gde je $\boldsymbol{X}_c = (X_c^{(x)}, X_c^{(y)}, \dots X_c^{(x)}, X_c^{(y)})$
  $(X_c^{(x)}, X_c^{(y)})$ je pozicija centra modela u koordinatnom sistemu slike.

ASM

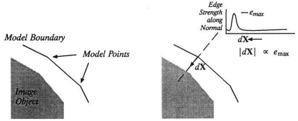
- Ako imamo procenjenu lokaciju objekta u slici (time i procenjenu skalu, rotaciju i translaciju) želimo da popravimo tu procenu.
- To znači da popravimo θ, s i $\boldsymbol{X}_c$ a tek zatim parametre b

ASM

- Ako imamo procenjenu lokaciju objekta u slici (time i procenjenu skalu, rotaciju i translaciju) želimo da popravimo tu procenu.
- To znači da popravimo θ, s i $\boldsymbol{X}_c$ a tek zatim parametre b
- Cilj je da pomerimo ključne tačke modela tako da se poklope sa ključnim tačkama slike koje se uglavnom nalaze na ivicama.

ASM

- Ako imamo procenjenu lokaciju objekta u slici (time i procenjenu skalu, rotaciju i translaciju) želimo da popravimo tu procenu.
- To znači da popravimo θ, s i $\boldsymbol{X}_c$ a tek zatim parametre b
- Cilj je da pomerimo ključne tačke modela tako da se poklope sa ključnim tačkama slike koje se uglavnom nalaze na ivicama.
- Jedan način je da povučemo normalu na ivicu modela i da pomerimo tačku modela ka najjačoj ivici slike (intenzitetom proporcionalnim jačini ivice)



ASM rescaling

- Ako je $\lambda_i = \sigma_i^2$ varijansa i-tog parametra vektora b i ne postoji korelacija parametara i i j za $i \neq j$ (što je tačno jer smo primenili PCA) tada je Mahalanobis distanca data sa $D_m = \sqrt{\sum_{i=1}^{K} \frac{b_i^2}{\lambda_i}}$

ASM rescaling

- Ako je $\lambda_i = \sigma_i^2$ varijansa i-tog parametra vektora b i ne postoji korelacija parametara i i j za $i \neq j$ (što je tačno jer smo primenili PCA) tada je Mahalanobis distanca data sa $D_m = \sqrt{\sum_{i=1}^{K} \frac{b_i^2}{\lambda_i}}$
- Oblik koji smo dobili kažemo da je valjan ako važi $D_m \leq D_{max}$ gde je $D_{max} = 3$ za 3 standardne devijacije.

ASM rescaling

- Ako je $\lambda_i = \sigma_i^2$ varijansa i-tog parametra vektora b i ne postoji korelacija parametara i i j za $i \neq j$ (što je tačno jer smo primenili PCA) tada je Mahalanobis distanca data sa $D_m = \sqrt{\sum_{i=1}^{K} \frac{b_i^2}{\lambda_i}}$
- Oblik koji smo dobili kažemo da je valjan ako važi $D_m \leq D_{max}$ gde je $D_{max} = 3$ za 3 standardne devijacije.
- Ako oblik nije valjan reskaliramo $b_i = b_i \frac{D_{max}}{D_m}$

Pregled predavanja

RLMS
Uvod

- Kao sto smo videli ASM traži sledećeg kandidata duž normale i odmah zatim pristupa podešavanju globalnih parametara. Ovo je previše pojednostavljeno i ne uzima u obzir okolinu.

RLMS
Uvod

- Kao sto smo videli ASM traži sledećeg kandidata duž normale i odmah zatim pristupa podešavanju globalnih parametara. Ovo je previše pojednostavljeno i ne uzima u obzir okolinu.
- ASM može da ne konvergira ka dobrom rešenju

RLMS
Uvod

- Kao sto smo videli ASM traži sledećeg kandidata duž normale i odmah zatim pristupa podešavanju globalnih parametara. Ovo je previše pojednostavljeno i ne uzima u obzir okolinu.
- ASM može da ne konvergira ka dobrom rešenju
- Takođe početna verzija ASM-a ne radi ništa po pitanu okluzija

RLMS
Uvod

- Kao sto smo videli ASM traži sledećeg kandidata duž normale i odmah zatim pristupa podešavanju globalnih parametara. Ovo je previše pojednostavljeno i ne uzima u obzir okolinu.
- ASM može da ne konvergira ka dobrom rešenju
- Takođe početna verzija ASM-a ne radi ništa po pitanu okluzija
- Rezultati nisu uvek najsrećniji

RLMS
Uvod

- Više nije 1992 pa možemo malo da uložimo u složeniji ali tačniji algoritam

RLMS
Uvod

- Više nije 1992 pa možemo malo da uložimo u složeniji ali tačniji algoritam
- Medicinske primene kao i primene gde se zahteva preciznost i doslednost traže bolje

RLMS
Uvod

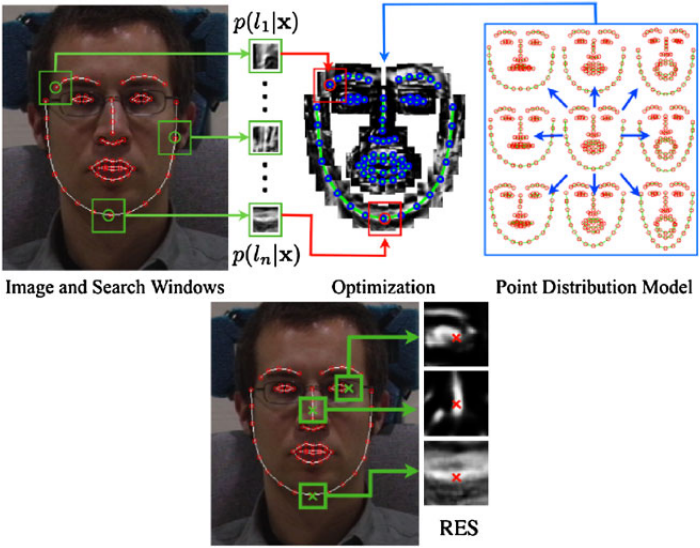
- Više nije 1992 pa možemo malo da uložimo u složeniji ali tačniji algoritam
- Medicinske primene kao i primene gde se zahteva preciznost i doslednost traže bolje
- 'Savitljivi modeli' se istražuju preko 2 decenije pa postoji veći broj algoritama koji se bave ovim problemom.

RLMS
Uvod

- Više nije 1992 pa možemo malo da uložimo u složeniji ali tačniji algoritam
- Medicinske primene kao i primene gde se zahteva preciznost i doslednost traže bolje
- 'Savitljivi modeli' se istražuju preko 2 decenije pa postoji veći broj algoritama koji se bave ovim problemom.
- Najviše obećavaju algoritmi koji nezavisno predviđaju lokacije ključnih tačaka modela, u kombinaciji sa nametnutim prior-om nad distribucijom PDM parametara.

RLMS

Uvod


$p(l_1|\mathbf{x})$
$p(l_n|\mathbf{x})$
Image and Search Windows
Optimization
Point Distribution Model
RES

Pregled predavanja

Problem

- Problem je isti kao i kod ASM-a
- $\boldsymbol{x}_i = s\boldsymbol{R}(\overline{\boldsymbol{x}}_i + \boldsymbol{\Phi}_i\boldsymbol{q}) + \boldsymbol{t}$

Problem

- Problem je isti kao i kod ASM-a
- $\boldsymbol{x}_i = s\boldsymbol{R}(\overline{\boldsymbol{x}}_i + \boldsymbol{\Phi}_i \boldsymbol{q}) + \boldsymbol{t}$
- $\boldsymbol{R}$ je matrica rotacije, s faktor globalnog skaliranja $\boldsymbol{t}$ translacija, $\overline{\boldsymbol{x}}_i$ prosečna lokacija i-tog PDM parametra u koordinatnom sistemu poravnatih objekata (reference frame).

Problem

- Problem je isti kao i kod ASM-a
- $\boldsymbol{x}_i = s\boldsymbol{R}(\overline{\boldsymbol{x}}_i + \boldsymbol{\Phi_i q}) + \boldsymbol{t}$
- $\boldsymbol{R}$ je matrica rotacije, s faktor globalnog skaliranja $\boldsymbol{t}$ translacija, $\overline{\boldsymbol{x}}_i$ prosečna lokacija i-tog PDM parametra u koordinatnom sistemu poravnatih objekata (reference frame).
- $\boldsymbol{\Phi_i}$ je podmatrica matrice baza varijacija $\boldsymbol{\Phi}$ (tj sopstvenih vektora dobijenih PCA metodom) koja se odnosi na i-u ključnu tačku

Problem

- Problem je isti kao i kod ASM-a
- $\boldsymbol{x}_i = s\boldsymbol{R}(\overline{\boldsymbol{x}}_i + \boldsymbol{\Phi_i q}) + \boldsymbol{t}$
- $\boldsymbol{R}$ je matrica rotacije, s faktor globalnog skaliranja $\boldsymbol{t}$ translacija, $\overline{\boldsymbol{x}}_i$ prosečna lokacija i-tog PDM parametra u koordinatnom sistemu poravnatih objekata (reference frame).
- $\boldsymbol{\Phi_i}$ je podmatrica matrice baza varijacija $\boldsymbol{\Phi}$ (tj sopstvenih vektora dobijenih PCA metodom) koja se odnosi na i-u ključnu tačku
- $\boldsymbol{q}$ je skup ne-rigidnih parametara

Problem

- Problem je isti kao i kod ASM-a
- $\boldsymbol{x}_i = s\boldsymbol{R}(\overline{\boldsymbol{x}}_i + \boldsymbol{\Phi_i q}) + \boldsymbol{t}$
- $\boldsymbol{R}$ je matrica rotacije, s faktor globalnog skaliranja $\boldsymbol{t}$ translacija, $\overline{\boldsymbol{x}}_i$ prosečna lokacija i-tog PDM parametra u koordinatnom sistemu poravnatih objekata (reference frame).
- $\boldsymbol{\Phi_i}$ je podmatrica matrice baza varijacija $\boldsymbol{\Phi}$ (tj sopstvenih vektora dobijenih PCA metodom) koja se odnosi na i-u ključnu tačku
- $\boldsymbol{q}$ je skup ne-rigidnih parametara
- Sa $\boldsymbol{p} = \{s, \boldsymbol{R}, \boldsymbol{t}, \boldsymbol{q}\}$ ćemo označiti skup svih PDM parametara

Problem

- Cilj je minimizacija $\mathcal{Q}$ po $\boldsymbol{p}$

$$\mathcal{Q}(\boldsymbol{p}) = \mathcal{R}(\boldsymbol{p}) + \sum_{i=1}^{n} \mathcal{D}_i(\boldsymbol{x_i}; \mathcal{I})$$

Problem

- Cilj je minimizacija $\mathcal{Q}$ po $\boldsymbol{p}$

$$\mathcal{Q}(\boldsymbol{p}) = \mathcal{R}(\boldsymbol{p}) + \sum_{i=1}^{n} \mathcal{D}_i(\boldsymbol{x_i}; \mathcal{I})$$

- $\mathcal{R}$ regularizacija - kažnjava kompleksne oblike

Problem

- Cilj je minimizacija $\mathcal{Q}$ po $\boldsymbol{p}$

$$\mathcal{Q}(\boldsymbol{p}) = \mathcal{R}(\boldsymbol{p}) + \sum_{i=1}^{n} \mathcal{D}_i(\boldsymbol{x_i}; \mathcal{I})$$

- $\mathcal{R}$ regularizacija - kažnjava kompleksne oblike
- $\mathcal{R}$ je često vezana za Gausovu raspodelu ili Gausov miks model

Problem

- Cilj je minimizacija $\mathcal{Q}$ po $\boldsymbol{p}$

$$\mathcal{Q}(\boldsymbol{p}) = \mathcal{R}(\boldsymbol{p}) + \sum_{i=1}^{n} \mathcal{D}_i(\boldsymbol{x_i}; \mathcal{I})$$

- $\mathcal{R}$ regularizacija - kažnjava kompleksne oblike
- $\mathcal{R}$ je često vezana za Gausovu raspodelu ili Gausov miks model
- $\mathcal{D}_i$ data term - meri odudaranje za i-tu ključnu tačku na lokaciji $\boldsymbol{x}_i$ na slici $\mathcal{I}$

Problem

- Cilj je minimizacija $\mathcal{Q}$ po $\boldsymbol{p}$

$$\mathcal{Q}(\boldsymbol{p}) = \mathcal{R}(\boldsymbol{p}) + \sum_{i=1}^{n} \mathcal{D}_i(\boldsymbol{x_i}; \mathcal{I})$$

- $\mathcal{R}$ regularizacija - kažnjava kompleksne oblike
- $\mathcal{R}$ je često vezana za Gausovu raspodelu ili Gausov miks model
- $\mathcal{D}_i$ data term - meri odudaranje za i-tu ključnu tačku na lokaciji $\boldsymbol{x}_i$ na slici $\mathcal{I}$
- Primeri za $\mathcal{D}_i$ su Mahalanobis distanca ili boosted Harr-like klasifikator

Izbor optimizatora

- $\mathcal{D}_i$ obično pokazuje veliku količinu šuma na prostornom domenu $\boldsymbol{x}_i$ pa lokalne determinističke optimizacione metode (npr. Njutnova metoda) nisu stabilne.

Izbor optimizatora

- $\mathcal{D}_i$ obično pokazuje veliku količinu šuma na prostornom domenu $\boldsymbol{x}_i$ pa lokalne determinističke optimizacione metode (npr. Njutnova metoda) nisu stabilne.
- Stohastičke metode (npr. Simplex) su otpornije na šum, ali je konvergencija spora (naročito kod velikog broja ključnih tačaka)

Pretraga ključnih tačaka

- $\mathcal{D}_i$ zavisi samo od prostornih koordinata ključne tačke

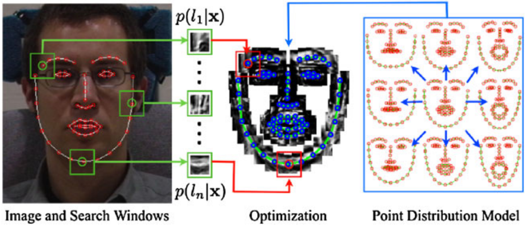
Pretraga ključnih tačaka

- $\mathcal{D}_i$ zavisi samo od prostornih koordinata ključne tačke
- Mozemo izvršiti lokalizovanu pretragu oko procenjene lokacije ključne tačke. Ta pretraga nam daje response mapu za datu ključnu tačku

Pretraga ključnih tačaka

- $\mathcal{D}_i$ zavisi samo od prostornih koordinata ključne tačke
- Mozemo izvršiti lokalizovanu pretragu oko procenjene lokacije ključne tačke. Ta pretraga nam daje response mapu za datu ključnu tačku
- Zatim radimo optimizaciju nad response mapama uzimajući u obzir distribuciju ključnih tačaka

Pretraga ključnih tačaka

- $\mathcal{D}_i$ zavisi samo od prostornih koordinata ključne tačke
- Mozemo izvršiti lokalizovanu pretragu oko procenjene lokacije ključne tačke. Ta pretraga nam daje response mapu za datu ključnu tačku
- Zatim radimo optimizaciju nad response mapama uzimajući u obzir distribuciju ključnih tačaka
- 

Likelihood, Prior

- Prior je pretpostavka raspodele PDM ključnih tačaka pre nego što uzmemo dokaze u obzir. Prior je u našem slučaju $p(\boldsymbol{p})$

Likelihood, Prior

- Prior je pretpostavka raspodele PDM ključnih tačaka pre nego što uzmemo dokaze u obzir. Prior je u našem slučaju $p(\boldsymbol{p})$
- Likelihood poravnatih ključnih tačaka pod pretpostavkom parametara modela je $p(\{\ell_i = 1\}_{i=1}^{n} | \{\boldsymbol{x}_i\}_{i=1}^{n}, \mathcal{I})$

Likelihood, Prior

- Prior je pretpostavka raspodele PDM ključnih tačaka pre nego što uzmemo dokaze u obzir. Prior je u našem slučaju $p(\boldsymbol{p})$
- Likelihood poravnatih ključnih tačaka pod pretpostavkom parametara modela je $p(\{\ell_i = 1\}_{i=1}^{n} | \{\boldsymbol{x}_i\}_{i=1}^{n}, \mathcal{I})$
- Posterior verovatnoća je verovatnoća parametara modela pod pretpostavkom da su sve ključne tačke poravnate tj. $p(\boldsymbol{p} | \{\ell_i = 1\}_{i=1}^{n}, \mathcal{I})$

Likelihood, Prior

- Prior je pretpostavka raspodele PDM ključnih tačaka pre nego što uzmemo dokaze u obzir. Prior je u našem slučaju $p(\boldsymbol{p})$
- Likelihood poravnatih ključnih tačaka pod pretpostavkom parametara modela je $p(\{\ell_i = 1\}_{i=1}^{n} | \{\boldsymbol{x}_i\}_{i=1}^{n}, \mathcal{I})$
- Posterior verovatnoća je verovatnoća parametara modela pod pretpostavkom da su sve ključne tačke poravnate tj. $p(\boldsymbol{p} | \{\ell_i = 1\}_{i=1}^{n}, \mathcal{I})$
- Evidence je $p(\{\ell_i = 1\}_{i=1}^{n} | \mathcal{I})$ - njega ćemo smatrati konstantom jer ne zavisi od parametara modela i nije bitan za optimizaciju.

Likelihood, Prior

- Prior je pretpostavka raspodele PDM ključnih tačaka pre nego što uzmemo dokaze u obzir. Prior je u našem slučaju $p(\boldsymbol{p})$
- Likelihood poravnatih ključnih tačaka pod pretpostavkom parametara modela je $p(\{\ell_i = 1\}_{i=1}^n | \{\boldsymbol{x}_i\}_{i=1}^n, \mathcal{I})$
- Posterior verovatnoća je verovatnoća parametara modela pod pretpostavkom da su sve ključne tačke poravnate tj. $p(\boldsymbol{p}|\{\ell_i = 1\}_{i=1}^n, \mathcal{I})$
- Evidence je $p(\{\ell_i = 1\}_{i=1}^n | \mathcal{I})$ - njega ćemo smatrati konstantom jer ne zavisi od parametara modela i nije bitan za optimizaciju.
- Primena Bayes-ove teoreme na Posterior i nezavisnost između detekcije ključnih tačaka nam daje

$$p(\boldsymbol{p}|\{\ell_i = 1\}_{i=1}^n, \mathcal{I}) \propto p(\boldsymbol{p}) \prod_{i=1}^n p(\ell_i = 1 | \boldsymbol{x}_i, \mathcal{I})$$

Patch Expert Uvod

- U slučaju da je prior uniforman prethodna formulacija je ekvivalentna sa Maximum Likelihood Estimate (ML) inače je ekvivalentan sa Maximum a-posterior estimate (MAP)

Patch Expert Uvod

- U slučaju da je prior uniforman prethodna formulacija je ekvivalentna sa Maximum Likelihood Estimate (ML) inače je ekvivalentan sa Maximum a-posterior estimate (MAP)
- Spajajući dve formulacije regularizacija dobija oblik: $\mathcal{R} = -\ln\{p(\boldsymbol{p})\}$
- A mera odudaranja i-te ključne tačke: $\mathcal{D}_i(\boldsymbol{x}_i; \mathcal{I}) = -\ln\{p(\ell_i = 1|\boldsymbol{x}_i, \mathcal{I})\}$

Patch Expert Uvod

- U slučaju da je prior uniforman prethodna formulacija je ekvivalentna sa Maximum Likelihood Estimate (ML) inače je ekvivalentan sa Maximum a-posterior estimate (MAP)
- Spajajući dve formulacije regularizacija dobija oblik: $\mathcal{R} = -\ln\{p(\boldsymbol{p})\}$
- A mera odudaranja i-te ključne tačke: $\mathcal{D}_i(\boldsymbol{x}_i; \mathcal{I}) = -\ln\{p(\ell_i = 1 | \boldsymbol{x}_i, \mathcal{I})\}$
- Likelihood poravnanja i-te ključne tačke modelujemo sa:

$$p(\ell_i | \boldsymbol{x}_i, \mathcal{I}) = \frac{1}{1 + e^{\ell_i \mathcal{C}_i(\boldsymbol{x}_i; \mathcal{I})}}$$

Patch Expert Uvod

- U slučaju da je prior uniforman prethodna formulacija je ekvivalentna sa Maximum Likelihood Estimate (ML) inače je ekvivalentan sa Maximum a-posterior estimate (MAP)
- Spajajući dve formulacije regularizacija dobija oblik: $\mathcal{R} = -\ln\{p(\boldsymbol{p})\}$
- A mera odudaranja i-te ključne tačke: $\mathcal{D}_i(\boldsymbol{x}_i; \mathcal{I}) = -\ln\{p(\ell_i = 1 | \boldsymbol{x}_i, \mathcal{I})\}$
- Likelihood poravnanja i-te ključne tačke modelujemo sa:

$$p(\ell_i | \boldsymbol{x}_i, \mathcal{I}) = \frac{1}{1 + e^{\ell_i \mathcal{C}_i(\boldsymbol{x}_i; \mathcal{I})}}$$

- $\mathcal{C}_i(\boldsymbol{x}_i; \mathcal{I})$ je logistic regresor i diskriminiše poravnate od neporavnatih lokacija i-te ključne tačke

Patch Expert problemi

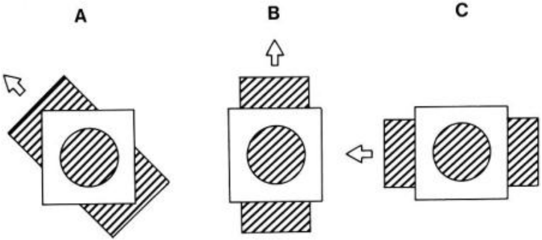
- Postojeće strategije optimizacije zahtevaju zamenu response mapa sa jednostavnijim parametrizovanim formama

Patch Expert problemi

- Postojeće strategije optimizacije zahtevaju zamenu response mapa sa jednostavnijim parametrizovanim formama
- U originalnom patch expertu ima previše šuma - nezahvalno za optimizaciju

Patch Expert problemi

- Postojeće strategije optimizacije zahtevaju zamenu response mapa sa jednostavnijim parametrizovanim formama
- U originalnom patch expertu ima previše šuma - nezahvalno za optimizaciju
- Detektori rade nad veoma malim oblastima pa su često neodređeni.
- Aperture problem



Prior modelovanje

- Pretpostavljamo da nerigidni parametri $\boldsymbol{q}$ vođeni Gausovom raspodelom

Prior modelovanje

- Pretpostavljamo da nerigidni parametri $\boldsymbol{q}$ vođeni Gausovom raspodelom
- $p(\boldsymbol{p}) \propto \mathcal{N}(\boldsymbol{q}; \mathbf{0}, \boldsymbol{\Lambda})$
  Gde je $\Lambda = diag\{[\lambda_1, \ldots, \lambda_m]\}$

Prior modelovanje

- Pretpostavljamo da nerigidni parametri $\boldsymbol{q}$ vođeni Gausovom raspodelom
- $p(\boldsymbol{p}) \propto \mathcal{N}(\boldsymbol{q}; \mathbf{0}, \boldsymbol{\Lambda})$
  Gde je $\Lambda = diag\{[\lambda_1, \ldots, \lambda_m]\}$
- Za rigidne parametre pretpostavljamo da su vođeni uniformnom raspodelom (tj. non-informative). Svaka translacija, rotacija, skaliranje je jednako verovatno.

Motivacije i problemi

Motivacije su nam:

- Da aproksimiramo response mapu funkcijom koja ima dovoljno lepa svojstva tako da je optimizacija efikasna i numerički stabilna

Motivacije i problemi

Motivacije su nam:

- Da aproksimiramo response mapu funkcijom koja ima dovoljno lepa svojstva tako da je optimizacija efikasna i numerički stabilna
- Da aproksimacija zadrži određenosti i neodređenosti originalne response mape.

Motivacije i problemi

Motivacije su nam:

- Da aproksimiramo response mapu funkcijom koja ima dovoljno lepa svojstva tako da je optimizacija efikasna i numerički stabilna
- Da aproksimacija zadrži određenosti i neodređenosti originalne response mape.

Problemi su:

- Kako da izbegnemo lokalne minimume.

Motivacije i problemi

Motivacije su nam:

- Da aproksimiramo response mapu funkcijom koja ima dovoljno lepa svojstva tako da je optimizacija efikasna i numerički stabilna
- Da aproksimacija zadrži određenosti i neodređenosti originalne response mape.

Problemi su:

- Kako da izbegnemo lokalne minimume.
- Kako da se izborimo sa lažnim detekcijama response mape.

Pregled predavanja

Aproksimacije response Mape
- Isotropic Gaussian Estimate
- Anisotropic Gaussian Estimate
- Gaussian Mixture Model Estimate
- Regularized landmark Mean-shift

Isotropic Gaussian Estimate

- Prvi način: Veoma sličan ASM modelu.
- Svaka ključna tačka PDM modela želi da se pomeri na optimalnu lokaciju. Te optimalne lokacije su u stvari maksimum response mape. (i-ta optimalna lokacija je $\boldsymbol{\mu}_i$)

Isotropic Gaussian Estimate

- Prvi način: Veoma sličan ASM modelu.
- Svaka ključna tačka PDM modela želi da se pomeri na optimalnu lokaciju. Te optimalne lokacije su u stvari maksimum response mape. (i-ta optimalna lokacija je $\boldsymbol{\mu}_i$)
- Cilj optimizacije minimizovati kvadrat razlike (uz razmatranje odgovarajućih težina ω_i i regularizationog dela)

$$\mathcal{Q}_{ISO} = \|\boldsymbol{q}\|^2_{\boldsymbol{\Lambda}^{-1}} + \sum_{i=1}^{n} \omega_i \|\boldsymbol{x}_i - \boldsymbol{\mu}_i\|^2$$

Isotropic Gaussian Estimate

- Prvi način: Veoma sličan ASM modelu.
- Svaka ključna tačka PDM modela želi da se pomeri na optimalnu lokaciju. Te optimalne lokacije su u stvari maksimum response mape. (i-ta optimalna lokacija je $\boldsymbol{\mu}_i$)
- Cilj optimizacije minimizovati kvadrat razlike (uz razmatranje odgovarajućih težina ω_i i regularizationog dela)

$$\mathcal{Q}_{ISO} = \|\boldsymbol{q}\|^2_{\boldsymbol{\Lambda}^{-1}} + \sum_{i=1}^{n} \omega_i \|\boldsymbol{x}_i - \boldsymbol{\mu}_i\|^2$$

- Težina ω_i odražava nivo pouzdanosti u koordinate našeg vrha ($\boldsymbol{\mu}_i$) i uglavnom je neka funkcija koja zavisi od $\{\boldsymbol{\mu}_i\}_{i=1}^{n}$

Probabilistička interpretacija

- Isto dobijamo ako modelujemo likelihood sa isotropskom Gausovom raspodelom.

Probabilistička interpretacija

- Isto dobijamo ako modelujemo likelihood sa isotropskom Gausovom raspodelom.
- $p(l_i = 1|\boldsymbol{x}_i, \mathcal{I}) = \frac{p(l_i=1, \boldsymbol{x}_i|\mathcal{I})p(l_i=1|\mathcal{I})}{p(\boldsymbol{x}_i|\mathcal{I})p(l_i=1|\mathcal{I})} = \frac{p(\boldsymbol{x}_i|l_i=1,\mathcal{I})p(l_i=1|\mathcal{I})}{p(\boldsymbol{x}_i|\mathcal{I})}$
  $\propto p(\boldsymbol{x}_i|l_i = 1, \mathcal{I}) \approx \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_i, \sigma_i^2 I)$

Probabilistička interpretacija

- Isto dobijamo ako modelujemo likelihood sa isotropskom Gausovom raspodelom.
- $p(l_i = 1|\boldsymbol{x}_i, \mathcal{I}) = \frac{p(l_i=1,\boldsymbol{x}_i|\mathcal{I})p(l_i=1|\mathcal{I})}{p(\boldsymbol{x}_i|\mathcal{I})p(l_i=1|\mathcal{I})} = \frac{p(\boldsymbol{x}_i|l_i=1,\mathcal{I})p(l_i=1|\mathcal{I})}{p(\boldsymbol{x}_i|\mathcal{I})}$ $\propto p(\boldsymbol{x}_i|l_i = 1, \mathcal{I}) \approx \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_i, \sigma_i^2 I)$
- Ako ubacimo dosadašnje formule u formulu posteriora: $-\ln p(\boldsymbol{p}) \prod_{i=1}^{n} p(\ell_i = 1|\boldsymbol{x}_i, \mathcal{I}) \approx$ $\|\boldsymbol{q}\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \|\boldsymbol{x}_i - \boldsymbol{\mu}_i\|^2$

Isotropic Gaussian Rešenje

- Tejlorov razvoj PDM parametara: $\boldsymbol{x}_i \approx \boldsymbol{x}_i^c + \boldsymbol{J}_i \Delta \boldsymbol{p}$

Isotropic Gaussian Rešenje

- Tejlorov razvoj PDM parametara: $\boldsymbol{x}_i \approx \boldsymbol{x}_i^c + \boldsymbol{J}_i \Delta \boldsymbol{p}$
- Rešavanje Gaus-Njutnovom metodom daje:

$$\Delta \boldsymbol{p} = -H_{ISO}^{-1}\big(\widetilde{\Lambda}^{-1}\boldsymbol{p} + \sum_{i=1}^{n} \omega_i \boldsymbol{J}_i^T (\boldsymbol{x}_i^c - \boldsymbol{\mu}_i)\big)$$

Isotropic Gaussian Rešenje

- Tejlorov razvoj PDM parametara: $\boldsymbol{x}_i \approx \boldsymbol{x}_i^c + \boldsymbol{J}_i \Delta \boldsymbol{p}$
- Rešavanje Gaus-Njutnovom metodom daje:

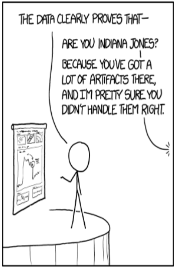
$$\Delta \boldsymbol{p} = -H_{ISO}^{-1}\left(\widetilde{\Lambda}^{-1}\boldsymbol{p} + \sum_{i=1}^{n} \omega_i \boldsymbol{J}_i^T (\boldsymbol{x}_i^c - \boldsymbol{\mu}_i)\right)$$

- H_{ISO} je aproksimacija Hesijana koja se računa kao: $H_{ISO} = \widetilde{\Lambda}^{-1} + \sum_{i=1}^{n} \omega_i \boldsymbol{J}_i^T \boldsymbol{J}_i$

Pregled predavanja

Anisotropic Gaussian Estimate Motivacija

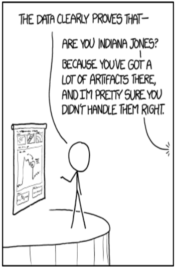
Isotropic Gausian Estimate Problemi:

- 

Očekujemo od logičkog regresora da je savršen (Da se maksimum response mape poklapa sa pravom lokacijom ključne tačke)

Anisotropic Gaussian Estimate Motivacija

Isotropic Gausian Estimate Problemi:

- 

  Očekujemo od logičkog regresora da je savršen (Da se maksimum response mape poklapa sa pravom lokacijom ključne tačke)

- Aperture problem. Iako jednostavan Isotropska Gausova procena daje nekad loše rezultate.

Anisotropic Gaussian Estimate

- Slično kao kod Isotropske Gausove procene samo aproksimiramo response mapu sa Gausovom raspodelom sa punom matricom kovarijanse.

$$p(\ell_i = 1 | \boldsymbol{x}_i, \mathcal{I}) \approx \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_i, \Sigma_i)$$

Anisotropic Gaussian Estimate

- Slično kao kod Isotropske Gausove procene samo aproksimiramo response mapu sa Gausovom raspodelom sa punom matricom kovarijanse.

$$p(\ell_i = 1|\boldsymbol{x}_i, \mathcal{I}) \approx \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_i, \Sigma_i)$$

- Postoji više metoda u zavisnosti od načina aproksimacije proseka i matrice kovarijanse.

Anisotropic Gaussian Estimate

- Slično kao kod Isotropske Gausove procene samo aproksimiramo response mapu sa Gausovom raspodelom sa punom matricom kovarijanse.

$$p(\ell_i = 1 | \boldsymbol{x}_i, \mathcal{I}) \approx \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_i, \Sigma_i)$$

- Postoji više metoda u zavisnosti od načina aproksimacije proseka i matrice kovarijanse.
- Jedan način je da se za prosek uzme maksimum originalne response mape a za matricu kovarijanse reši sledeći problem (mašinskim učenjem)

$$\Sigma_i = \sum_{\boldsymbol{x} \in \Psi_i} \frac{p(\ell_i = 1 | \boldsymbol{x}, \mathcal{I})}{\sum_{\boldsymbol{y} \in \Psi_i} p(\ell_i = 1 | \boldsymbol{y}, \mathcal{I})} (\boldsymbol{x} - \boldsymbol{\mu}_i)(\boldsymbol{x} - \boldsymbol{\mu}_i)^T$$

ANI

- Σ_i može da se posmatra i kao weighted biased sample covariance.

ANI

- Σ_i može da se posmatra i kao weighted biased sample covariance.
- Ubacivanjem aproksimacija u posterior dobijamo $\mathcal{Q}_{ANI} = \|q\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{\mu}_i\|^2_{\Sigma_i^{-1}}$

ANI

- Σ_i može da se posmatra i kao weighted biased sample covariance.
- Ubacivanjem aproksimacija u posterior dobijamo $\mathcal{Q}_{ANI} = \|q\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{\mu}_i\|^2_{\Sigma_i^{-1}}$
- Slično rešenje Gaus-Njutnovom metodom:

$$\Delta \boldsymbol{p} = -H_{ANI}^{-1}\left(\widetilde{\Lambda}^{-1} + \sum_{i=1}^{n} \boldsymbol{J}_i^T \Sigma_i^{-1}(\boldsymbol{x}_i^c - \boldsymbol{\mu}_i)\right)$$

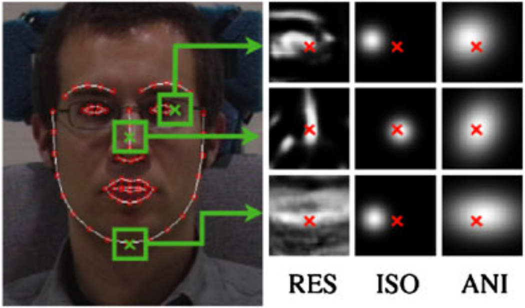
ANI

- Σ_i može da se posmatra i kao weighted biased sample covariance.
- Ubacivanjem aproksimacija u posterior dobijamo $\mathcal{Q}_{ANI} = \|q\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{\mu}_i\|^2_{\Sigma_i^{-1}}$
- Slično rešenje Gaus-Njutnovom metodom:

$$\Delta \boldsymbol{p} = -H_{ANI}^{-1}\left(\widetilde{\Lambda}^{-1} + \sum_{i=1}^{n} \boldsymbol{J}_i^T \Sigma_i^{-1} (\boldsymbol{x}_i^c - \boldsymbol{\mu}_i)\right)$$

- H_{ANI} je aproksimacija Hesijana: $H_{ANI} = \widetilde{\Lambda}^{-1} + \sum_{i=1}^{n} \boldsymbol{J}_i^T \Sigma_i^{-1} \boldsymbol{J}_i$

Primer



Pregled predavanja

Aproksimacije response Mape

- Isotropic Gaussian Estimate
- Anisotropic Gaussian Estimate
- Gaussian Mixture Model Estimate
- Regularized landmark Mean-shift

GMM estimate

- Response mapu aproksimiramo kao linearnu superpoziciju gausovih raspodela

$$p(\ell_i = 1|\boldsymbol{x}_i, \mathcal{I}) \approx \sum_{k=1}^{K_i} \pi_{ik} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_{ik}, \Sigma_{ik})$$

GMM estimate

- Response mapu aproksimiramo kao linearnu superpoziciju gausovih raspodela

$$p(\ell_i = 1|\boldsymbol{x}_i, \mathcal{I}) \approx \sum_{k=1}^{K_i} \pi_{ik} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_{ik}, \Sigma_{ik})$$

- K_i je broj Gausovih raspodela koji se kombinuju linearno pri aproksimaciji response mape i-te ključne tačke

GMM estimate

- Response mapu aproksimiramo kao linearnu superpoziciju gausovih raspodela

$$p(\ell_i = 1|\boldsymbol{x}_i, \mathcal{I}) \approx \sum_{k=1}^{K_i} \pi_{ik} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_{ik}, \Sigma_{ik})$$

- K_i je broj Gausovih raspodela koji se kombinuju linearno pri aproksimaciji response mape i-te ključne tačke
- π_{ik} je koeficijent linearne kombinacije

GMM estimate

- Ubacivanjem aproksimacije response mape u posterior dobijamo

$$p(\boldsymbol{p}|\{\ell_i\}_{i=1}^n, \mathcal{I}) \propto p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{k=1}^{K_i} p_i(z_{ik} = 1, \ell_i | \boldsymbol{x}_i, \mathcal{I})$$

- Maximum likelihood rešnje se nalazi pomoću EM algoritma

- Maximum likelihood rešnje se nalazi pomoću EM algoritma
- E korak sadrži računanje odgovornosti tj. posterior distribucije nad latentnim promenljivim: $E[z_{ik}] = p(z_{ik} = 1 | \ell_i = 1, \boldsymbol{x}_i, \mathcal{I})$

- M korak se sastoji iz minimizacije:

$$-\ln p(\boldsymbol{p}|\{\ell_i\}_{i=1}^{n}, \mathcal{I}) = -\ln p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{k=1}^{K_i} \pi_{ik} \mathcal{N}(\boldsymbol{x}_i; \mu_{ik}, \Sigma_{ik})$$

- M korak se sastoji iz minimizacije:

$$-\ln p(\boldsymbol{p}|\{\ell_i\}_{i=1}^{n}, \mathcal{I}) = -\ln p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{k=1}^{K_i} \pi_{ik} \mathcal{N}(\boldsymbol{x}_i; \mu_{ik}, \Sigma_{ik})$$

Što je isto kao i minimizacija očekivanja:
$GMM(p) = E_{\boldsymbol{z}}[-\ln p(\boldsymbol{p}) \prod_{i=1}^{n} p(\ell_i = 1, z_i|\boldsymbol{x}_i, \mathcal{I})] \propto$
$\|q\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \sum_{k=1}^{K_i} E[z_{ik}] \|\boldsymbol{x}_i - \mu_{ik}|^2_{\Sigma_{ik}}$

- Rešenje se dobija iterativno Gauss-Njutnovom metodom optimizacije

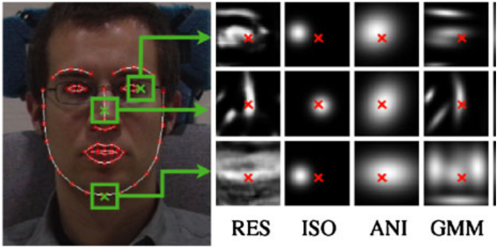
- Rešenje se dobija iterativno Gauss-Njutnovom metodom optimizacije
- $$\Delta p = -H_{GMM}^{-1}(\widetilde{\Lambda}^{-1} p + \sum_{i=1}^{n} \sum_{k=1}^{K_i} \omega_{ik} \boldsymbol{J}_i^T \Sigma_{ik}^{-1} (\boldsymbol{x}_i^c - \mu_{ik}))$$

- Rešenje se dobija iterativno Gauss-Njutnovom metodom optimizacije
-

$$\Delta p = -H_{GMM}^{-1}(\widetilde{\Lambda}^{-1}p + \sum_{i=1}^{n}\sum_{k=1}^{K_i}\omega_{ik}\boldsymbol{J}_i^T\Sigma_{ik}^{-1}(\boldsymbol{x}_i^c - \mu_{ik}))$$

- H_{GMM} aproksimacija hesijana:

$$\widetilde{\Lambda}^{-1} + \sum_{i=1}^{n}\sum_{k=1}^{K_i}\omega_{ik}\boldsymbol{J}_i^T\Sigma_{ik}^{-1}\boldsymbol{J}_i$$


RES
ISO
ANI
GMM

GMM Problemi

- Proces procene GMM parametara iz response mape je nelinearna optimizacija sama za sebe.

GMM Problemi

- Proces procene GMM parametara iz response mape je nelinearna optimizacija sama za sebe.
- Fitovanje GMM parametara radimo za SVAKU ključnu tačku!

GMM Problemi

- Proces procene GMM parametara iz response mape je nelinearna optimizacija sama za sebe.
- Fitovanje GMM parametara radimo za SVAKU ključnu tačku!
- Lokalno konvergentna i zahteva da se fiksira broj gausovih raspodela za svaku ključnu tačku.

GMM Problemi

- Proces procene GMM parametara iz response mape je nelinearna optimizacija sama za sebe.
- Fitovanje GMM parametara radimo za SVAKU ključnu tačku!
- Lokalno konvergentna i zahteva da se fiksira broj gausovih raspodela za svaku ključnu tačku.
- Aproksimacije za proseke i kovarijanse uglavnom nisu optimalne i dolaze po cenu lošije aproksimacije response mape

GMM Problemi

- Proces procene GMM parametara iz response mape je nelinearna optimizacija sama za sebe.
- Fitovanje GMM parametara radimo za SVAKU ključnu tačku!
- Lokalno konvergentna i zahteva da se fiksira broj gausovih raspodela za svaku ključnu tačku.
- Aproksimacije za proseke i kovarijanse uglavnom nisu optimalne i dolaze po cenu lošije aproksimacije response mape
- Multimodalne raspodele (uvek) $\implies$ uglavnom ćemo završiti u lokalnom optimumu

Pregled predavanja

RLMS

- Kada smo primenili PCA odsekli smo određeni broj vektora, dakle nikad ne možemo tačno rekonstruistati lokacije ključnih tačaka

RLMS

- Kada smo primenili PCA odsekli smo određeni broj vektora, dakle nikad ne možemo tačno rekonstruistati lokacije ključnih tačaka
- Ako smo zadržali dovoljan stepen varijabilnosti za ostatak možemo da smatramo da potiče iz šuma

RLMS

- Kada smo primenili PCA odsekli smo određeni broj vektora, dakle nikad ne možemo tačno rekonstruistati lokacije ključnih tačaka
- Ako smo zadržali dovoljan stepen varijabilnosti za ostatak možemo da smatramo da potiče iz šuma
- Modelujemo šum kao homoscedastičke izotropske Gausove raspodele:

$$\boldsymbol{y}_i = \boldsymbol{x}_i + \boldsymbol{\epsilon}_i$$

Gde $\boldsymbol{\epsilon}_i \approx \mathcal{N}(\boldsymbol{\epsilon}_i; \mathbf{0}, \rho I)$

RLMS

- Kada smo primenili PCA odsekli smo određeni broj vektora, dakle nikad ne možemo tačno rekonstruistati lokacije ključnih tačaka
- Ako smo zadržali dovoljan stepen varijabilnosti za ostatak možemo da smatramo da potiče iz šuma
- Modelujemo šum kao homoscedastičke izotropske Gausove raspodele:

$$\boldsymbol{y}_i = \boldsymbol{x}_i + \boldsymbol{\epsilon}_i$$

  Gde $\boldsymbol{\epsilon}_i \approx \mathcal{N}(\boldsymbol{\epsilon}_i; \mathbf{0}, \rho I)$
- ρ je varijansa šuma lokacija ključnih tačaka

- Moghaddam & Pentland 1997:

$$\rho = \frac{1}{N-m} \sum_{i=m+1}^{N} \lambda_i$$

- Moghaddam & Pentland 1997:

$$\rho = \frac{1}{N-m} \sum_{i=m+1}^{N} \lambda_i$$

- Označimo skup kandidata za i-tu ključnu tačku sa Ψ_i

- Moghaddam & Pentland 1997:

$$\rho = \frac{1}{N-m} \sum_{i=m+1}^{N} \lambda_i$$

- Označimo skup kandidata za i-tu ključnu tačku sa Ψ_i
- Ako posmatramo lokaciju prave ključne tačke kao latentnu promenljivu možemo da nađemo likelihood iz marginalizacije:

$$p(\ell_i = 1 | \boldsymbol{x}_i, \mathcal{I}) = \sum_{\boldsymbol{y}_i \in \Psi_i} p(\ell_i = 1 | \boldsymbol{y}_i, \mathcal{I}) p(\boldsymbol{y}_i | \boldsymbol{x}_i)$$

- Moghaddam & Pentland 1997:

$$\rho = \frac{1}{N - m} \sum_{i=m+1}^{N} \lambda_i$$

- Označimo skup kandidata za i-tu ključnu tačku sa Ψ_i
- Ako posmatramo lokaciju prave ključne tačke kao latentnu promenljivu možemo da nađemo likelihood iz marginalizacije:

$$p(\ell_i = 1 | \boldsymbol{x}_i, \mathcal{I}) = \sum_{\boldsymbol{y}_i \in \Psi_i} p(\ell_i = 1 | \boldsymbol{y}_i, \mathcal{I}) p(\boldsymbol{y}_i | \boldsymbol{x}_i)$$

- $p(\boldsymbol{y}_i | \boldsymbol{x}_i) = \mathcal{N}(\boldsymbol{y}_i | \boldsymbol{x}_i, \rho I)$ i $p(\ell_i = 1 | \boldsymbol{y}_i, \mathcal{I}) = \pi_{\boldsymbol{y}_i}$

- Do sada smo aproksimirali response mapu parametrizovanim oblikom, sada ćemo aproksimirati neparametarskom procenom.

- Do sada smo aproksimirali response mapu parametrizovanim oblikom, sada ćemo aproksimirati neparametarskom procenom.
- Gaussian Kernel Density Estimate:

$$p(\ell_i = 1 | \boldsymbol{x}_i, \mathcal{I}) = \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{y}_i | \boldsymbol{x}_i, \rho I)$$

- Do sada smo aproksimirali response mapu parametrizovanim oblikom, sada ćemo aproksimirati neparametarskom procenom.
- Gaussian Kernel Density Estimate:

$$p(\ell_i = 1|\boldsymbol{x}_i, \mathcal{I}) = \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{y}_i|\boldsymbol{x}_i, \rho I)$$

- Kvalitet procene zavisi isključivo od kvaliteta uzoraka $\boldsymbol{y}_i$

- Do sada smo aproksimirali response mapu parametrizovanim oblikom, sada ćemo aproksimirati neparametarskom procenom.
- Gaussian Kernel Density Estimate:

$$p(\ell_i = 1|\boldsymbol{x}_i, \mathcal{I}) = \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{y}_i|\boldsymbol{x}_i, \rho \boldsymbol{I})$$

- Kvalitet procene zavisi isključivo od kvaliteta uzoraka $\boldsymbol{y}_i$
- Ako uzorkujemo previše retko procena parametra ρ će biti manja nego što je potrebno.

- Kao i ranije

$$p(\boldsymbol{p}|\{\ell_i = 1\}_{i=1}^{n}, \mathcal{I}) \propto p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{y}_i | \boldsymbol{x}_i, \rho \boldsymbol{I})$$

- Kao i ranije

$$p(\boldsymbol{p}|\{\ell_i = 1\}_{i=1}^{n}, \mathcal{I}) \propto p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{y}_i | \boldsymbol{x}_i, \rho \boldsymbol{I})$$

- Kao i kod GMM-a maksimizujemo EM algoritmom

- Kao i ranije

$$p(\boldsymbol{p}|\{\ell_i = 1\}_{i=1}^{n}, \mathcal{I}) \propto p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{y}_i | \boldsymbol{x}_i, \rho I)$$

- Kao i kod GMM-a maksimizujemo EM algoritmom
- E korak - (računamo posterior distribuciju nad latentnim promenljivim):

$$E[z_{i\boldsymbol{y}_i}] = p(\boldsymbol{y}_i | \ell_i = 1, \boldsymbol{x}_i, \mathcal{I})$$

- M korak - maksimizacija EM algoritmom:

$$p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho \boldsymbol{I})$$

- M korak - maksimizacija EM algoritmom:

$$p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho I)$$

- Isto što minimizacija očekivanja:
$E_{\boldsymbol{z}}[-\ln p(\boldsymbol{p}) \prod_{i=1}^{n} \prod_{\boldsymbol{y}_i \in \Psi_i} (\pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho I))^{z_{i\boldsymbol{y}_i}}]$

- M korak - maksimizacija EM algoritmom:

$$p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho I)$$

- Isto što minimizacija očekivanja: $E_{\boldsymbol{z}}[-\ln p(\boldsymbol{p}) \prod_{i=1}^{n} \prod_{\boldsymbol{y}_i \in \Psi_i} (\pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho I))^{z_{i\boldsymbol{y}_i}}]$
-
$$\propto \|\boldsymbol{q}\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \omega_{\boldsymbol{y}_i} \frac{1}{\rho} \|\boldsymbol{x}_i - \boldsymbol{y}_i\|^2$$

- Rešenje je:

$$\Delta p = -H_{KDE}^{-1}(\widetilde{\Lambda}^{-1} p + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i} \omega_{\boldsymbol{y}_i} \frac{1}{\rho} \boldsymbol{J}_i^T (\boldsymbol{x}_i^c - \boldsymbol{y}_i))$$

- Rešenje je:

$$\Delta p = -H_{KDE}^{-1}(\widetilde{\Lambda}^{-1} p + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i} \omega_{\boldsymbol{y}_i} \frac{1}{\rho} \boldsymbol{J}_i^T (\boldsymbol{x}_i^c - \boldsymbol{y}_i))$$

- H_{KDE} aproksimacija hesijana:

$$\widetilde{\Lambda}^{-1} + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i} \frac{\omega_{\boldsymbol{y}_i}}{\rho} \boldsymbol{J}_i^T \boldsymbol{J}_i$$

- Sređivanjem dobijemo:

$$\Delta p = -(\rho\widetilde{\Lambda}^{-1} + \boldsymbol{J}^T\boldsymbol{J})(\rho\widetilde{\Lambda}^{-1}p - \boldsymbol{J}^T\boldsymbol{v})$$

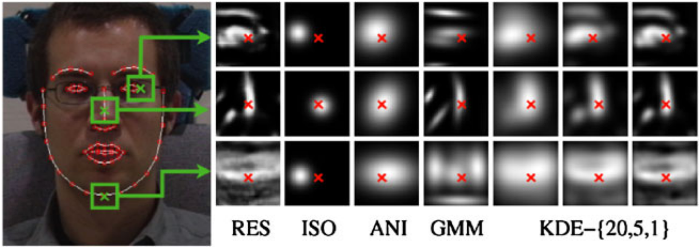
- Sređivanjem dobijemo:

$$\Delta p = -(\rho\widetilde{\Lambda}^{-1} + \boldsymbol{J}^T\boldsymbol{J})(\rho\widetilde{\Lambda}^{-1}p - \boldsymbol{J}^T\boldsymbol{v})$$

- Gde je $\boldsymbol{v} = [\boldsymbol{v}_1, \ldots \boldsymbol{v}_n]^T$ i

$$\boldsymbol{v}_i = \sum_{\boldsymbol{y}_i \in \Psi_i} \omega_{\boldsymbol{y}_i}\boldsymbol{y}_i - \boldsymbol{x}_i^c = \sum_{\boldsymbol{y}_i \in \Psi_i} \left(\frac{\pi_{\boldsymbol{y}_i}\mathcal{N}(\boldsymbol{x}_i^c; \boldsymbol{y}_i, \rho I)}{\sum_{\boldsymbol{z}_i \in \Psi_i} \pi_{\boldsymbol{z}_i}\mathcal{N}(\boldsymbol{x}_i^c; \boldsymbol{z}_i, \rho I)}\boldsymbol{y}_i\right) - \boldsymbol{x}_i^c$$

- $\boldsymbol{v}_i$ je mean-shift vektor za i-tu ključnu tačku


RES
ISO
ANI
GMM
KDE–{20,5,1}

Algorithm 1: Regularized landmark mean-shift optimization algorithm

Input: $\mathcal{I}$, $\boldsymbol{p}$

Output: $\boldsymbol{p}$

function RLMS ($\mathcal{I}, \boldsymbol{p}$)

Compute responses $p(\ell_i | \boldsymbol{x}_i, \mathcal{I}) = \frac{1}{1+e^{\ell_i C_i(\boldsymbol{x}_i;\mathcal{I})}}$

while *not_converged($\boldsymbol{p}$)* **do**

  Linearize $\boldsymbol{x}_i \approx \boldsymbol{x}_i^c + \boldsymbol{J}_i \Delta p$

  Compute mean-shift vektors

  $$\boldsymbol{v}_i = \sum_{\boldsymbol{y}_i \in \Psi_i} \left(\frac{\pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i^c; \boldsymbol{y}_i, \rho \boldsymbol{I})}{\sum_{\boldsymbol{z}_i \in \Psi_i} \pi_{\boldsymbol{z}_i} \mathcal{N}(\boldsymbol{x}_i^c; \boldsymbol{z}_i, \rho \boldsymbol{I})} \boldsymbol{y}_i \right) - \boldsymbol{x}_i^c$$

  Compute PDM parameter update:

  $$\Delta p = -(\rho \widetilde{\Lambda}^{-1} + \boldsymbol{J}^T \boldsymbol{J})(\rho \widetilde{\Lambda}^{-1} p - \boldsymbol{J}^T \boldsymbol{v})$$

  Update parameters: $\boldsymbol{p} = \boldsymbol{p} + \Delta \boldsymbol{p}$

end

return $\boldsymbol{p}$

Izbegavanje lokalnih optimuma

- ρ kao varijansa kernela može da se posmatra kao parametar glatkosti.

Izbegavanje lokalnih optimuma

- ρ kao varijansa kernela može da se posmatra kao parametar glatkosti.
- Carreira-Perpinan-Willims 2003: Postoji $\rho < \infty$ tako da je KDE unimodalna bez obzira na distribuciju uzoraka.

Izbegavanje lokalnih optimuma

- ρ kao varijansa kernela može da se posmatra kao parametar glatkosti.
- Carreira-Perpinan-Willims 2003: Postoji $\rho < \infty$ tako da je KDE unimodalna bez obzira na distribuciju uzoraka.
- Povećavanje ρ možemo da posmatramo kao da radimo glađu procenu aproksimirane response mape za malo ρ.

Izbegavanje lokalnih optimuma

- ρ kao varijansa kernela može da se posmatra kao parametar glatkosti.
- Carreira-Perpinan-Willims 2003: Postoji $\rho < \infty$ tako da je KDE unimodalna bez obzira na distribuciju uzoraka.
- Povećavanje ρ možemo da posmatramo kao da radimo glađu procenu aproksimirane response mape za malo ρ.
- Krenemo od velikog ρ nađemo željeni mod zatim smanjimo ρ

Izbegavanje lokalnih optimuma

- ρ kao varijansa kernela može da se posmatra kao parametar glatkosti.
- Carreira-Perpinan-Willims 2003: Postoji $\rho < \infty$ tako da je KDE unimodalna bez obzira na distribuciju uzoraka.
- Povećavanje ρ možemo da posmatramo kao da radimo glađu procenu aproksimirane response mape za malo ρ.
- Krenemo od velikog ρ nađemo željeni mod zatim smanjimo ρ

Rukovanje smetnjama

- Umesto kvadratne funckije koristimo M-estimator (Geman-McClure): $\rho(r^2; \theta) = \frac{r^2}{r^2+\theta^2}$

Rukovanje smetnjama

- Umesto kvadratne funckije koristimo M-estimator (Geman-McClure): $\rho(r^2;\theta) = \frac{r^2}{r^2+\theta^2}$
- Optimizujemo $\|\boldsymbol{q}\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \varpi_{\boldsymbol{y}_i} \rho(\|\boldsymbol{x}_i - \boldsymbol{y}_i\|^2;\theta)$

Rukovanje smetnjama

- Umesto kvadratne funckije koristimo M-estimator (Geman-McClure): $\rho(r^2;\theta) = \frac{r^2}{r^2+\theta^2}$
- Optimizujemo $\|\boldsymbol{q}\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \varpi_{\boldsymbol{y}_i} \rho(\|\boldsymbol{x}_i - \boldsymbol{y}_i\|^2;\theta)$
- Moramo da prilagodimo težine $\varpi_{\boldsymbol{y}_i}$ novoj parametrizaciji $p(\boldsymbol{y}_i|\boldsymbol{x}_i)$

Rukovanje smetnjama

- Umesto kvadratne funckije koristimo M-estimator (Geman-McClure): $\rho(r^2;\theta) = \frac{r^2}{r^2+\theta^2}$
- Optimizujemo $\|\boldsymbol{q}\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \varpi_{\boldsymbol{y}_i} \rho(\|\boldsymbol{x}_i - \boldsymbol{y}_i\|^2;\theta)$
- Moramo da prilagodimo težine $\varpi_{\boldsymbol{y}_i}$ novoj parametrizaciji $p(\boldsymbol{y}_i|\boldsymbol{x}_i)$
- $p(\boldsymbol{y}_i|\boldsymbol{x}_i) = \frac{1}{\mathcal{Z}(\theta)} e^{-\rho(\|\boldsymbol{x}_i - \boldsymbol{y}_i\|^2;\theta)}$

Rukovanje smetnjama

- Umesto kvadratne funckije koristimo M-estimator (Geman-McClure): $\rho(r^2;\theta) = \frac{r^2}{r^2+\theta^2}$
- Optimizujemo $\|\boldsymbol{q}\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \varpi_{\boldsymbol{y}_i} \rho(\|\boldsymbol{x}_i - \boldsymbol{y}_i\|^2;\theta)$
- Moramo da prilagodimo težine $\varpi_{\boldsymbol{y}_i}$ novoj parametrizaciji $p(\boldsymbol{y}_i|\boldsymbol{x}_i)$
- $p(\boldsymbol{y}_i|\boldsymbol{x}_i) = \frac{1}{\mathcal{Z}(\theta)} e^{-\rho(\|\boldsymbol{x}_i-\boldsymbol{y}_i\|^2;\theta)}$
- Posterior nad latentnim računamo kao: $\varpi_{\boldsymbol{y}_i} = \frac{\pi_{\boldsymbol{y}_i} e^{-\rho(\|\boldsymbol{x}_i-\boldsymbol{y}_i\|^2;\theta)}}{\sum_{\boldsymbol{z}_i \in \Psi_i} \pi_{\boldsymbol{z}_i} e^{-\rho(\|\boldsymbol{x}_i-\boldsymbol{z}_i\|^2;\theta)}}$

Eksperimenti - Slike

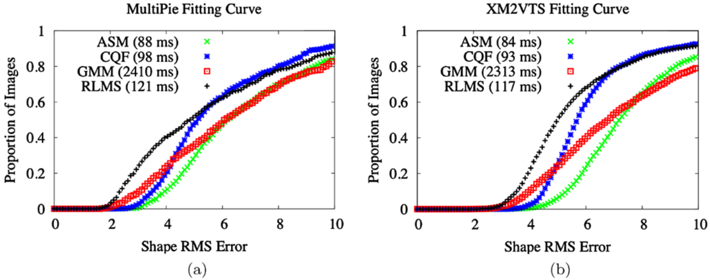


Figure: (a) MULTIPIE 68 ključnih tačaka 762 slike lica 339 subjekata (b)XM2VTS 68 ključnih tačaka 2360 slika lica sa 295 subjekata

Eksperimenti - Video

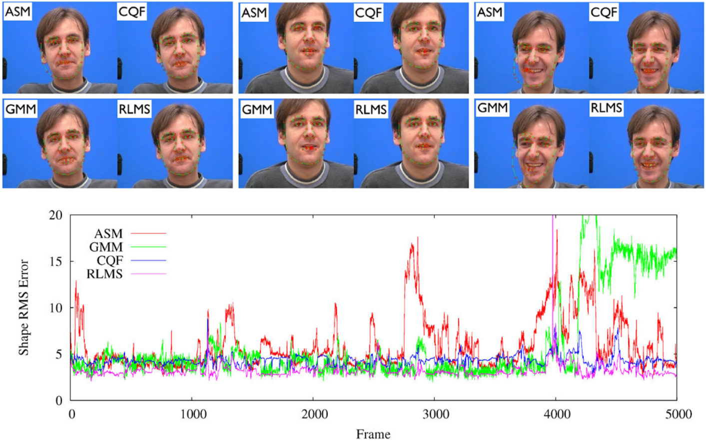


Figure: FGNet Talking Face baza. Prikazani su rezultati za slike {0,1230,4200}

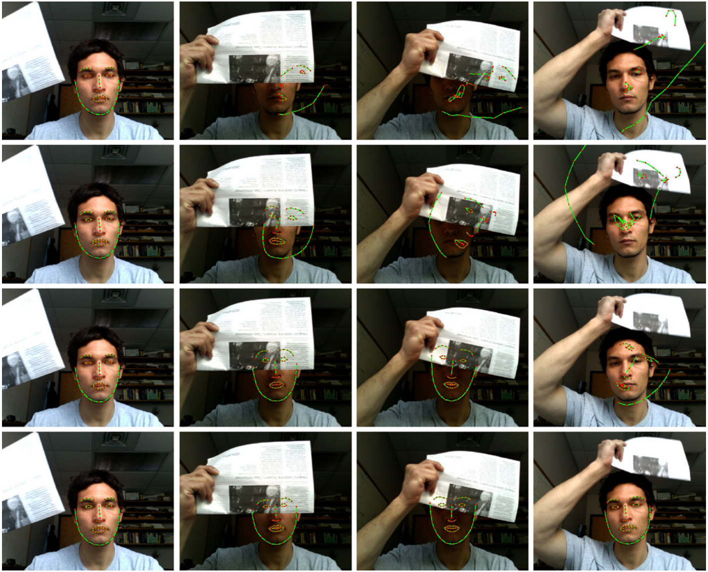
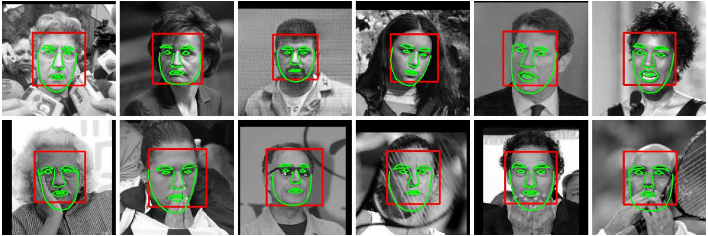


Figure: Od vrha na dole: ML + Gaussian, MAP + Gaussian, ML + Geman McClure, MAP + Geman McClure

Figure: Od vrha na dole: ML + Gaussian, MAP + Gaussian, ML + Geman McClure, MAP + Geman McClure

In the Wild



Pitanja

Pregled predavanja

Formule za Data Manupulation

Skaliramo i rotiramo ostale oblike tako da se poravnaju sa z_1.
Skaliramo oblik j za s_j i rotiramo za ugao θ_j.
$s_j = \sqrt{a_j^2 + b_j^2}$, $\theta_j = \tan^{-1} \frac{b_j}{a_j}$
Gde su $a_j = \frac{z_1 \cdot z_j}{\|z_j\|^2}$ i $b_j = \frac{\sum_{i=1}^{n} z_{ji}^{(x)} z_{1i}^{(y)} - z_{1i}^{(x)} z_{ji}^{(y)}}{\|z_j\|^2}$

Nove tačke dobijamo kao

$$\begin{bmatrix} z_{ji}^{\hat{(x)}} \\ z_{ji}^{\hat{(y)}} \end{bmatrix} = s_j \begin{bmatrix} \cos\theta_j & \sin\theta_j \\ -\sin\theta_j & \cos\theta_j \end{bmatrix} \begin{bmatrix} z_{ji}^{(x)} \\ z_{ji}^{(y)} \end{bmatrix}$$

Pregled predavanja

PCA Formule

- Računamo prosek: $\mu = \frac{1}{S}\sum_{i=1}^{S} z_i$ gde je S veličina trening skupa.
- Računamo matricu kovarijanse: $\Sigma = \frac{1}{S-1}\sum_{i=1}^{S}(z_j - \mu)(z_j - \mu)^T$
- Tražimo sopstvene vrednosti i sopstvene vektore matrice Σ. Neka su to λ_1,..λ_S i v_1 ,.. v_S poređani po veličini sopstvenih vrednosti.
- Biramo K tako da zadržimo određeni procenat varijabilnosti koji računamo kao $\frac{\sum_{i=1}^{K}\lambda_i}{\sum_{j=1}^{S}\lambda_j}$

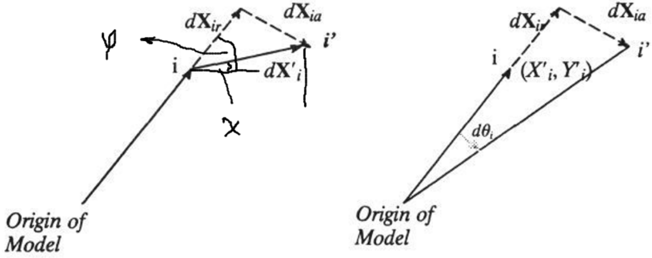
Pregled predavanja

ASM Iterativni metod

- Ako je $\boldsymbol{X}$ trenutni PDM objekta a $\boldsymbol{X} + d\boldsymbol{X}$ pomereni PDM trebamo naći $d\theta$, ds i $d\boldsymbol{X}_c$ a zatim i db tako da dođe do poklapanja
- Translacija $d\boldsymbol{X}_c = (dX_c^{(x)}, dX_c^{(y)}, \dots dX_c^{(x)}, dX_c^{(y)})$: $dX_c^{(x)} = \frac{1}{n}\sum_{i=1}^{n} dX_i^{(x)}$ i $dX_c^{(y)} = \frac{1}{n}\sum_{i}^{n} dX_i^{(y)}$
- $dX'^{(x)}_i = dX_i^{(x)} - dX_c^{(x)}$ i $dX'^{(y)}_i = dX_i^{(y)} - dX_c^{(y)}$ $\boldsymbol{X}' = \boldsymbol{X} - \boldsymbol{X}_c$
- Sada tražimo rotaciju i skaliranje tako da $\boldsymbol{X}'$ se poklopi što je bolje sa $d\boldsymbol{X}' + \boldsymbol{X}'$

ASM Iterativni metod

- Skaliranje $ds_i = \frac{|dX_{ir}|}{|X'|}$ i $ds \approx \frac{1}{n}\sum_{i=1}^{n} ds_i$



- dX_{ir} dobijamo iz $\cos(\phi - \chi) = \frac{|dX_{ir}|}{\sqrt{dX'^2_i + dY'^2_i}}$
- $d\theta \approx \frac{1}{n}\sum_{i=1}^{n} d\theta_i$ gde $d\theta_i \approx \frac{|dX_{ia}|}{|X'_i + dX_{ir}|}$

ASM Iterativni metod

- Ostalo nam je još da izračunamo promenu parametara db i ako je $x + dx \approx \mu + P(b + db)$ tada je $db \approx P^T dx$
- Uglavnom se parametri ažuriraju sa određenim težinama
- $X_c = X_c + \omega_t dX_c$, $\theta = \theta + \omega_\theta d\theta$ i $s = s(1 + \omega_s ds)$
- Za promenu parametara b uglavnom koristimo dijagonalnu matricu sa težinama proporcionalnim standardnim devijacijama parametara.
- $b = b + W_b db$
- Poslednji korak je da analiziramo da su parametri b unutar zadovoljavajućih granica.

ASM drugi način

- Postavimo $b = 0$
- Izračunamo $x = \mu + Pb$
- Nađemo θ, s, i translaciju $d\boldsymbol{X}_c$ tako da $\boldsymbol{X} + d\boldsymbol{X}$ najbolje pomerimo na $\boldsymbol{X}$ - Procrustes algoritam za $\boldsymbol{X}$ i $\boldsymbol{X} + d\boldsymbol{X}$
- Primenom translacije , rotacije i skaliranjem dobili smo Y koji zatim projektujemo na naš prostor poravnatih objekata - čime dobijemo y - (Procrustes algoritam sa z_1)
- Azuriramo parametre modela $b = P^T(y - \mu)$ i idemo na drugi korak do konvergencije

Pregled predavanja

Patch Expert

- $\mathcal{C}_i(\boldsymbol{x}_i; \mathcal{I}) = \boldsymbol{w}_i \mathcal{P}(\mathcal{W}(\boldsymbol{x}_i; \mathcal{I})) + b_i$ gde su $\boldsymbol{w}_i$ gain i b_i bias.
- $\mathcal{P}(\boldsymbol{r})$ normalizuje $\boldsymbol{r}$ na prosek 0 i varijansu 1 .
- $\mathcal{W}(\boldsymbol{x}_i; \mathcal{I})$ predstavlja kvadratni deo slike $\mathcal{I}$ centriran u $\boldsymbol{x}_i$
- Prednost ovakvog klasifikatora je što response map-u možemo da računamo pomoću konvolucije (što je efikasno)

Pregled predavanja

Isotropic Equivalency

Ako ubacimo dosadašnje formule u formulu posteriora:

$$-\ln p(\boldsymbol{p})\prod_{i=1}^{n} p(\ell_i = 1|\boldsymbol{x}_i, \mathcal{I}) = -\ln \frac{1}{(2\pi)^{\frac{m}{2}}(\lambda_1 \ldots \lambda_m)^{\frac{1}{2}}} + \frac{1}{2}\boldsymbol{q}^T\Lambda^{-1}\boldsymbol{q} - \sum_{i=1}^{n}(\ln \frac{1}{2\pi\sigma_i^2} - \frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_i)^T(\sigma_i^2 I)^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_i)$$

$$= \frac{1}{2}\|\boldsymbol{q}\|^2_{\Lambda^{-1}} + \frac{1}{2}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\|\boldsymbol{x}_i - \boldsymbol{\mu}_i\|^2 + \sum_{i=1}^{n}(\ln 2\pi\sigma_i^2) + \ln (2\pi)^{\frac{m}{2}}(\lambda_1 \ldots \lambda_m)^{\frac{1}{2}}$$

Pregled predavanja

Joint Probability

- Važi:

$$p_i(z_{ik}=1,\ell_i|\boldsymbol{x}_i,\mathcal{I}) = p(z_{ik}=1|\boldsymbol{x}_i,\mathcal{I})p(\ell_i=1|z_{ik}=1,\boldsymbol{x}_i,\mathcal{I})$$

$$= \pi_{ik}\mathcal{N}(\boldsymbol{x}_i;\mu_{ik},\Sigma_{ik})$$

- $p(z_{ik}=1|\boldsymbol{x}_i,\mathcal{I}) = \pi_{ik}$
- $p(\ell_i=1|z_{ik}=1,\boldsymbol{x}_i,\mathcal{I}) = \mathcal{N}(\boldsymbol{x}_i;\mu_{ik},\Sigma_{ik})$

Pregled predavanja

GMM Responsibility

E korak sadrži računanje odgovornosti tj. posterior distribucije nad latentnim promenljivim:

$$E[z_{ik}] = p(z_{ik} = 1|\ell_i = 1, \boldsymbol{x}_i, \mathcal{I}) = \frac{p(z_{ik}=1,\ell_i=1|\boldsymbol{x}_i,\mathcal{I})}{\sum_{j=1}^{K_i} p(z_{ij}=1,\ell_i=1|\boldsymbol{x}_i,\mathcal{I})} = \frac{p(\ell_i=1|z_{ik}=1,\boldsymbol{x}_i,\mathcal{I})p(z_{ik}=1|\boldsymbol{x}_i,\mathcal{I})}{\sum_{j=1}^{K_i} p(\ell_i=1|z_{ij}=1,\boldsymbol{x}_i,\mathcal{I})p(z_{ij}=1|\boldsymbol{x}_i,\mathcal{I})} = \frac{\pi_{ik}\mathcal{N}(\boldsymbol{x}_i;\mu_{ik},\Sigma_{ik})}{\sum_{j=1}^{K_i} \pi_{ij}\mathcal{N}(\boldsymbol{x}_i;\mu_{ij},\Sigma_{ij})}$$

Pregled predavanja

GMM M Step

M korak se sastoji iz minimizacije:

$$-\ln p(\boldsymbol{p} | \{\ell_i\}_{i=1}^{n}, \mathcal{I}) = -\ln p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{k=1}^{K_i} \pi_{ik} \mathcal{N}(\boldsymbol{x}_i; \mu_{ik}, \Sigma_{ik})$$

GMM M Step

M korak se sastoji iz minimizacije:

$$-\ln p(\boldsymbol{p}|\{\ell_i\}_{i=1}^n, \mathcal{I}) = -\ln p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{k=1}^{K_i} \pi_{ik} \mathcal{N}(\boldsymbol{x}_i; \mu_{ik}, \Sigma_{ik})$$

Što je isto kao i minimizacija očekivanja:
$GMM(p) = E_{\boldsymbol{z}}[-\ln p(\boldsymbol{p}) \prod_{i=1}^{n} p(\ell_i = 1, z_i|\boldsymbol{x}_i, \mathcal{I})] =$

GMM M Step

M korak se sastoji iz minimizacije:

$$-\ln p(\boldsymbol{p}|\{\ell_i\}_{i=1}^n, \mathcal{I}) = -\ln p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{k=1}^{K_i} \pi_{ik} \mathcal{N}(\boldsymbol{x}_i; \mu_{ik}, \Sigma_{ik})$$

Što je isto kao i minimizacija očekivanja:
$GMM(p) = E_{\boldsymbol{z}}[-\ln p(\boldsymbol{p}) \prod_{i=1}^n p(\ell_i = 1, z_i|\boldsymbol{x}_i, \mathcal{I})] =$

$E_{\boldsymbol{z}}[-\ln p(\boldsymbol{p}) \prod_{i=1}^n \prod_{k=1}^{K_i} (\pi_{ik} \mathcal{N}(\boldsymbol{x}_i; \mu_{ik}, \Sigma_{ik})^{z_{ik}})] \propto$
$\|q\|_{\Lambda^{-1}}^2 + \sum_{i=1}^n \sum_{k=1}^{K_i} E[z_{ik}] \|\boldsymbol{x}_i - \mu_{ik}|_{\Sigma_{ik}}^2$

Pregled predavanja

Gauss Mixture Model(GMM)

- Primene: Data mining, pattern recognition, statistical analysis, machine learning.
- Koristeći dovoljan broj Gausovih raspodela i podešavajući proseke kao i matrice kovarijanse skoro svaka neprekidna raspodela može da se aproksimira sa proizvoljnom tačnošću pomoću GMM
- Superpozicija K gausovih raspodela oblika:

$$p(x) = \sum_{k=1}^{K} \mathcal{N} \pi_k(x|\mu_k, \Sigma_k)$$

Se naziva gausov miks model

- Integracijom po x dobijamo: $\sum_{k=1}^{K} \pi_k = 1$
  Dok $p(x) \geq 0 \wedge \mathcal{N}(x; \mu_k, \Sigma_k) \geq 0 \implies \pi_k \geq 0$
- Uvedimo K dimenzionu binarnu promenljivu $\boldsymbol{z}$ tako da je $\sum_{k=1}^{K} z_k = 1$.
- Tada definišemo zajedničku raspodelu $p(\boldsymbol{x}, \boldsymbol{z})$ preko marginalne $p(\boldsymbol{z})$ i uslovne $p(\boldsymbol{x}|\boldsymbol{z})$
- Marginalnu raspodelu $p(\boldsymbol{z})$ izvlačimo iz činjenice da je $p(z_k = 1) = \pi_k$ odakle sledi $p(\boldsymbol{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$
- $p(\boldsymbol{x}|\boldsymbol{z}) = \prod_{k=1}^{K} \mathcal{N}(x; \mu_k, \Sigma_k)^{z_k}$

- $p(\boldsymbol{x}) = \sum_{\boldsymbol{z}} p(\boldsymbol{x}, \boldsymbol{z})$ odakle sledi da znamo da predstavimo GMM kao marginalizaciju zajedničke raspodele koja uključuje latentne promenljive.
- Još jedna bitna mera je odgovornost komponente k pri generisanju $\boldsymbol{x}$:

$$\gamma(z_k) = p(z_k = 1|\boldsymbol{x}) = \frac{p(z_k = 1)p(\boldsymbol{x}|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(\boldsymbol{x}|z_j = 1)}$$

$$= \frac{\pi_k \mathcal{N}(\boldsymbol{x}|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\boldsymbol{x}|\mu_j, \Sigma_j)}$$

- Ako probamo da dobijemo vrednosti $\boldsymbol{\pi}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$ iz MLE $p(\boldsymbol{x}|\boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \prod_{k=1}^{K} p(x_i|\boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$
  Odakle sledi
  $\ln p(\boldsymbol{x}|\boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \sum_{i=1}^{N} \ln \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k)$

- Postavljamo $\frac{\partial \ln p(\boldsymbol{x}|\boldsymbol{\pi},\boldsymbol{\Sigma},\boldsymbol{\mu})}{\partial \mu_k} = 0$ i $\frac{\partial \ln p(\boldsymbol{x}|\boldsymbol{\pi},\boldsymbol{\Sigma},\boldsymbol{\mu})}{\partial \Sigma_k} = 0$
- Odatle sledi

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) x_n$$

Gde $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$ i $\gamma(z_{nk}) = \gamma(z_k = 1|x_n)$

- $$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T$$

- Želimo da maksimizujemo $\ln p(\boldsymbol{x}|\boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ tako da $\sum_{k=1}^{K} \pi_k = 1$
Možemo uvesti lagranžev koeficijent zatim maksimizovati

- Iz $\frac{\partial \ln p(\boldsymbol{x}|\boldsymbol{\pi},\boldsymbol{\Sigma},\boldsymbol{\mu}) + \lambda(\sum_{k=1}^{K} \pi_k - 1)}{\partial pi_k} = 0$ sledi $\sum_{n=1}^{N} \gamma(z_{nk}) + \lambda \pi_k = 0$
- Kada prosumiramo za vrednosti k dobijamo $\lambda = -N$ Odatle sledi $\pi_k = \frac{N_k}{N}$
- Ovo nije zatvorena forma rešenja jer $\gamma(z_{nk})$ zavisi od π_k
- Motivacija za iterativni Expecation Maximization algoritam (EM)

- Kako znamo da predstavimo GMM marginalizacijom zajedničke raspodele ima smisla videti da li možemo rešiti problem maksimizacije likelihood-a kompletnog skupa podataka $\{X, Z\}$
- $$p(\boldsymbol{x}, \boldsymbol{z} | \boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k))^{z_{nk}}$$

  Odatle sledi $\ln p(\boldsymbol{x}, \boldsymbol{z} | \boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} (\ln \pi_k + \ln \mathcal{N}(x_n; \mu_k, \Sigma_k))$

- Postavljanjem $\frac{\partial \ln p(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{\pi},\boldsymbol{\Sigma},\boldsymbol{\mu})}{\partial \mu_l} = 0$ i $\frac{\partial \ln p(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{\pi},\boldsymbol{\Sigma},\boldsymbol{\mu})}{\partial \Sigma_l} = 0$ dobijamo

$$\mu_l = \frac{1}{\sum_{n=1}^{N} z_{nl}} \sum_{n=1}^{N} z_{nl} x_n$$

$$\Sigma_l = \frac{1}{\sum_{n=1}^{N} z_{nl}} \sum_{n=1}^{N} z_{nl} (x_n - \mu_l)(x_n - \mu_l)^T$$

- Slično kao i pre $\lambda = -N \implies \pi_k = \frac{1}{N} \sum_{n=1}^{N} z_{nk}$

- Kako u praksi nemamo date vrednosti latentnih promenljivih zelimo da nađemo njihova očekivanja $E[z_n k]$
- Za to nam treba $p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \frac{p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) p(\boldsymbol{z}|\boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu})}{p(\boldsymbol{x}|\boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu})} \propto p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) p(\boldsymbol{z}|\boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$
- $p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) \propto \prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k))^{z_{nk}}$

- $E[z_{nk}] = \sum_{z_{nk}} z_{nk} \frac{(\pi_k \mathcal{N}(x_n;\mu_k,\Sigma_k))^{z_{nk}}}{\sum_{z_{nl}} \pi_l \mathcal{N}(x_n;\mu_l,\Sigma_l)} = \frac{\pi_k \mathcal{N}(x_n;\mu_k,\Sigma_k)}{\sum_{z_{nl}} (\pi_l \mathcal{N}(x_n;\mu_l,\Sigma_l))^{z_{nl}}} = \frac{\pi_k \mathcal{N}(x_n;\mu_k,\Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n;\mu_j,\Sigma_j)} = \gamma(z_{nk}) = \gamma(z_k = 1|x_n)$
- Očekivana vrednost log likelihood funkcije sa kompletnim podacima je

$$E_{\boldsymbol{z}}[\ln p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\pi}, \boldsymbol{\Sigma}, \boldsymbol{\mu})] = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk})(\ln \pi_k + \ln \mathcal{N}(x_n; \mu_k, \Sigma_k))$$

EM Algoritam

- Biramo inicijalne vrednosti za $\boldsymbol{\mu}^{old}$, $\boldsymbol{\Sigma}^{old}$ i $\boldsymbol{\pi}^{old}$
- Evaluiramo odgovornosti $E[z_{nk}] = \gamma(z_{nk})$
- Koristimo $\gamma(z_{nk})$ umesto z_{nk} pri računanju $\boldsymbol{\mu}^{new}$, $\boldsymbol{\Sigma}^{new}$ i $\boldsymbol{\pi}^{new}$
- Evaluiramo očekivanu vrednost log likelihood funkcije sa kompletnim podacima
- Vraćamo se na korak 2 dok ne postignemo konvergenciju

Pregled predavanja

RLMS Responsibility

E korak - (računamo posterior distribuciju nad latentnim promenljivim):

$$E[z_{i\boldsymbol{y}_i}] = p(\boldsymbol{y}_i|\ell_i = 1, \boldsymbol{x}_i, \mathcal{I}) = \frac{\pi_{\boldsymbol{y}_i}\mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho I)}{\sum_{\boldsymbol{z}_i \in \Psi_i} \pi_{\boldsymbol{z}_i}\mathcal{N}(\boldsymbol{x}_i; \boldsymbol{z}_i, \rho I)}$$

Pregled predavanja

RLMS M Step

- M korak - maksimizacija EM algoritmom:

$$p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho \boldsymbol{I})$$

RLMS M Step

- M korak - maksimizacija EM algoritmom:

$$p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho I)$$

- Isto što minimizacija očekivanja:
  $\mathcal{Q}_{KDE}(\boldsymbol{p}) =$
  $E_{\boldsymbol{z}}[-\ln p(\boldsymbol{p}) \prod_{i=1}^{n} \prod_{\boldsymbol{y}_i \in \Psi_i} (\pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho I))^{z_{i\boldsymbol{y}_i}}]$

RLMS M Step

- M korak - maksimizacija EM algoritmom:

$$p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho I)$$

- Isto što minimizacija očekivanja:
$\mathcal{Q}_{KDE}(\boldsymbol{p}) =$
$E_{\boldsymbol{z}}[-\ln p(\boldsymbol{p}) \prod_{i=1}^{n} \prod_{\boldsymbol{y}_i \in \Psi_i} (\pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho I))^{z_{i\boldsymbol{y}_i}}]$
$\propto \frac{1}{2} \|\boldsymbol{q}\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} E[z_{i\boldsymbol{y}_i}](-\ln \pi_{\boldsymbol{y}_i} + \frac{D}{2} \ln 2\pi +$
$\frac{1}{2} \ln \rho I + \frac{1}{2\rho} \|\boldsymbol{x}_i - \boldsymbol{y}_i\|^2)$

RLMS M Step

- M korak - maksimizacija EM algoritmom:

$$p(\boldsymbol{p}) \prod_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho I)$$

- Isto što minimizacija očekivanja:
  $\mathcal{Q}_{KDE}(\boldsymbol{p}) =$
  $E_{\boldsymbol{z}}[-\ln p(\boldsymbol{p}) \prod_{i=1}^{n} \prod_{\boldsymbol{y}_i \in \Psi_i} (\pi_{\boldsymbol{y}_i} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{y}_i, \rho I))^{z_{i\boldsymbol{y}_i}}]$
  $\propto \frac{1}{2} \|\boldsymbol{q}\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} E[z_{i\boldsymbol{y}_i}](-\ln \pi_{\boldsymbol{y}_i} + \frac{D}{2} \ln 2\pi +$
  $\frac{1}{2} \ln \rho I + \frac{1}{2\rho} \|\boldsymbol{x}_i - \boldsymbol{y}_i\|^2)$
- $$\propto \|\boldsymbol{q}\|^2_{\Lambda^{-1}} + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i \in \Psi_i} \omega_{\boldsymbol{y}_i} \frac{1}{\rho} \|\boldsymbol{x}_i - \boldsymbol{y}_i\|^2$$

Pregled predavanja

Appendix RLMS Step

- Rešenje je:

$$\Delta p = -H_{KDE}^{-1}(\widetilde{\Lambda}^{-1}p + \sum_{i=1}^{n}\sum_{\boldsymbol{y}_i}\omega_{\boldsymbol{y}_i}\frac{1}{\rho}\boldsymbol{J}_i^T(\boldsymbol{x}_i^c - \boldsymbol{y}_i))$$

Appendix RLMS Step

- Rešenje je:

$$\Delta p = -H_{KDE}^{-1}(\widetilde{\Lambda}^{-1} p + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i} \omega_{\boldsymbol{y}_i} \frac{1}{\rho} \boldsymbol{J}_i^T (\boldsymbol{x}_i^c - \boldsymbol{y}_i))$$

- H_{KDE} aproksimacija hesijana:

$$\widetilde{\Lambda}^{-1} + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i} \frac{\omega_{\boldsymbol{y}_i}}{\rho} \boldsymbol{J}_i^T \boldsymbol{J}_i$$

Appendix RLMS Step

- Rešenje je:

$$\Delta p = -H_{KDE}^{-1}(\widetilde{\Lambda}^{-1} p + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i} \omega_{\boldsymbol{y}_i} \frac{1}{\rho} \boldsymbol{J}_i^T (\boldsymbol{x}_i^c - \boldsymbol{y}_i))$$

- H_{KDE} aproksimacija hesijana:

$$\widetilde{\Lambda}^{-1} + \sum_{i=1}^{n} \sum_{\boldsymbol{y}_i} \frac{\omega_{\boldsymbol{y}_i}}{\rho} \boldsymbol{J}_i^T \boldsymbol{J}_i$$

- $H_{KDE} = \frac{1}{\rho}(\rho\widetilde{\Lambda}^{-1} + \sum_{i=1}^{n} \boldsymbol{J}_i^T \boldsymbol{J}_i \sum_{\boldsymbol{y}_i} \omega_{\boldsymbol{y}_i}) = \frac{1}{\rho}(\rho\widetilde{\Lambda}^{-1} + \boldsymbol{J}^T \boldsymbol{J})$

Appendix RLMS Step

- Rešenje je:

$$\Delta p = -H_{KDE}^{-1}(\widetilde{\Lambda}^{-1}p + \sum_{i=1}^{n}\sum_{\boldsymbol{y}_i}\omega_{\boldsymbol{y}_i}\frac{1}{\rho}\boldsymbol{J}_i^T(\boldsymbol{x}_i^c - \boldsymbol{y}_i))$$

- H_{KDE} aproksimacija hesijana:

$$\widetilde{\Lambda}^{-1} + \sum_{i=1}^{n}\sum_{\boldsymbol{y}_i}\frac{\omega_{\boldsymbol{y}_i}}{\rho}\boldsymbol{J}_i^T\boldsymbol{J}_i$$

- $H_{KDE} = \frac{1}{\rho}(\rho\widetilde{\Lambda}^{-1} + \sum_{i=1}^{n}\boldsymbol{J}_i^T\boldsymbol{J}_i\sum_{\boldsymbol{y}_i}\omega_{\boldsymbol{y}_i}) = \frac{1}{\rho}(\rho\widetilde{\Lambda}^{-1} + \boldsymbol{J}^T\boldsymbol{J})$
- $\implies \Delta p = -(\rho\widetilde{\Lambda}^{-1} + \boldsymbol{J}^T\boldsymbol{J})^{-1}(\rho\widetilde{\Lambda}^{-1}p - \sum_{i=1}^{n}\boldsymbol{J}_i^T(\sum_{\boldsymbol{y}_i}\omega_{\boldsymbol{y}_i}\boldsymbol{y}_i - \boldsymbol{x}_i\sum_{\boldsymbol{y}_i}\omega_{\boldsymbol{y}_i}))$

Literatura I

Christopher M. Bishop
Pattern Recognition and Machine Learning
2006.

Jason M. Saragih, Simon Lucey, Jeffrey F. Cohn
Deformable Model Fitting by Regularized Landmark Mean-Shift
September 2010

T.F. Cootes, J.F. Taylor
Active Shape Models - 'Smart Snakes'
1992.