Regresiona Gausova uslovna slučajna polja

Mladen Nikolić

Grupa za mašinsko učenje i primene
Matematički fakultet
Univerzitet u Beogradu

Pregled

CCRF

GCRF

FF-GCRF

Pregled

CCRF

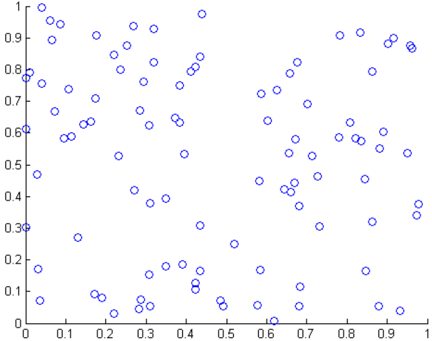
GCRF

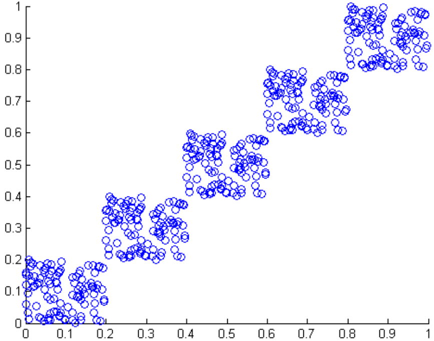
FF-GCRF

Uslovna nezavisnost promenljivih

- Promenljive y_1 i y_2 su uslovno nezavisne pri uslovu x ukoliko važi $P(y_1, y_2|x) = P(y_1|x)P(y_2|x)$
- Ovo je slabija relacija od pune nezavisnosti

Uslovna nezavisnost promenljivih





Uslovna nezavisnost promenljivih

- Uobičajena pretpostavka mašinskog učenja je da su opažanja nezavisna i jednako raspodeljena (eng. IID)
- Ova nezavisnost se prilikom modelovanja ogleda u činjenici da je model $y = f(x)$ funkcija vektora atributa x i da time za neko konkretno x_i, y_i zavisi samo od tog x_i, ali ne i od drugih y_j za $j \neq i$
- Odnosno, $P(y_1, \ldots, y_n | x_1, \ldots, x_n) = \prod_{i=1}^{n} P(y_i | x_i)$
- Koliko je ovo realistično?

Realističnost nezavisnosti promenljivih

- Promene temperature vazduha
- Rasprostiranje supstanci u zemljištu
- Kupovina proizvoda
- Citiranje naučnih radova

Uslovna slučajna polja (CRF)

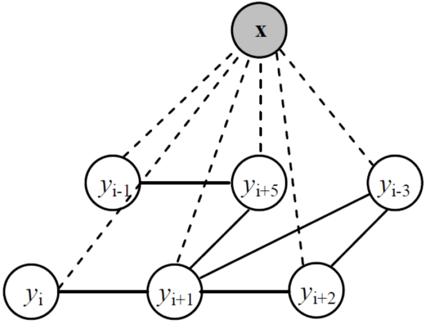
- Diskriminativni model
- Modeluju $P(y_1, \ldots, y_n | x_1 \ldots, x_n)$ ne pretpostavljajući faktorizaciju na pojedinačne verovatnoće, već u skladu sa grafom zavisnosti između promenljivih
- Polazne formulacije su bile klasifikacione, ali postoje i regresione

Neprekidna uslovna slučajna polja (CCRF)

- Neprekidna uslovna slučajna polja:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})} \exp\left(\sum_{i=1}^{N} A(\boldsymbol{\alpha}, y_i, \mathbf{x}) + \sum_{i \sim j} I(\boldsymbol{\beta}, y_i, y_j, \mathbf{x})\right)$$

gde $i \sim j$ označava povezanost promenljivih y_i i y_j u grafu zavisnosti



Neprekidna uslovna slučajna polja (CCRF)

- Predviđanje:

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}} P(\mathbf{y}|\mathbf{x})$$

- Učenje:

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \log P(\mathbf{y}|\mathbf{x})$$

$$(\hat{\alpha}, \hat{\beta}) = \arg\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} L(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

pri čemu je moguće dodati i regularizaciju po parametrima $\boldsymbol{\alpha}$ i $\boldsymbol{\beta}$

Atributske funkcije

- Kako izabrati funkcije A i I?

$$A(\boldsymbol{\alpha}, y_i, \mathbf{x}) = \sum_{k=1}^{K} \alpha_k f_k(y_i, \mathbf{x})$$

$$I(\boldsymbol{\beta}, y_i, y_j, \mathbf{x}) = \sum_{l=1}^{L} \beta_l g_l(y_i, y_j, \mathbf{x})$$

- Za proizvoljne izbore atributskih funkcija f_k i g_k, trening i predviđanje mogu biti računski vrlo zahtevni

Pregled

CCRF

GCRF

FF-GCRF

Gausova uslovna slučajna polja (GCRF)

- Ukoliko su f_k i g_k kvadratne funkcije, $P(\mathbf{y}|\mathbf{x})$ je višedimenzionalna Gausova raspodela!
- Gausova raspodela se lako ocenjuje, a predviđanje se svodi na izračunavanje proseka
- Kako izabrati funkcije f_k i g_k na koristan način?

Izbor atributskih funkcija

- Neka su dati modeli $R_{ik}(\mathbf{x})$, $k = 1, \ldots, K$ modeli koji predviđaju y_i za $i = 1, \ldots, N$ i grafovi zavisnosti G_l, $l = 1, \ldots, L$
- Onda možemo definisati

$$f_k(y_i, \mathbf{x}) = -(y_i - R_{ik}(\mathbf{x}))^2, \; k = 1, \ldots, K$$

$$g_l(y_i, y_j, \mathbf{x}) = -e^l_{ij} S^l_{ij}(\mathbf{x})(y_i - y_j)^2$$

pri čemu važi

$$e^l_{ij} = \begin{cases} 1, & (i,j) \in G_l \\ 0, & \text{inače} \end{cases}$$

Gausova uslovna slučajna polja

- Puna forma:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})} \exp\left(-\sum_{i=1}^{N}\sum_{k=1}^{K}\alpha_k(y_i - R_{ik}(\mathbf{x}))^2 - \sum_{i\sim j}\sum_{l=1}^{L}\beta_l e_{ij}^l S_{ij}^l(\mathbf{x})(y_i - y_j)^2\right)$$

- Pojednostavljena forma:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x}, \alpha, \beta)} \exp\left(-\sum_{i=1}^{N}\alpha(y_i - R_i(x_i))^2 - \sum_{i\sim j}\beta S_{ij}(y_i - y_j)^2\right)$$

Gausova forma

- Treba upariti:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})} \exp\left(-\sum_{i=1}^{N}\sum_{k=1}^{K} \alpha_k (y_i - R_{ik}(\mathbf{x}))^2 - \sum_{i\sim j}\sum_{l=1}^{L} \beta_l e_{ij}^l S_{ij}^l(\mathbf{x})(y_i - y_j)^2 \right)$$

i

$$-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}) = -\frac{1}{2}\mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} + \mathbf{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + const$$

Gausova forma

$$\boldsymbol{\Sigma}_{ij}^{-1} = \begin{cases} 2\sum_{k=1}^{K}\alpha_k + 2\sum_{k=1}^{K}\sum_{l=1}^{L}\beta_l e_{ik}^l S_{ik}^l(\mathbf{x}) & i = j \\ -2\sum_{l=1}^{L}\beta_l e_{ij}^l S_{ij}^l & i \neq j \end{cases}$$

$$\boldsymbol{\mu} = \boldsymbol{\Sigma}\mathbf{b}$$

$$b_i = 2\left(\sum_{k=1}^{K}\alpha_k R_{ik}(\mathbf{x})\right)$$

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$$

Gausova forma

- Da li je $\mathbf{\Sigma}^{-1}$ pozitivno semi-definitna matrica?
- Da bi bila, dovoljno je da bude dijagonalno dominantna
- To važi ako je $\alpha_k > 0$ i $\beta_l > 0$ za svako k i l
- Ovim se metod ograničava!
- Postoji unapređenje koje otklanja ovo ograničenje

Predviđanje

- Predviđanje se daje prosekom Gausove raspodele

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}) = \mathbf{\Sigma b}$$

Predviđanje

- Predviđanje se daje prosekom Gausove raspodele

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}) = \mathbf{\Sigma b}$$

- Za predviđanje je potrebno uraditi inverziju matrice $\mathbf{\Sigma}^{-1}$, što je računski skuplje nego predviđanja klasičnih metoda

Učenje

- Potrebno je izračunati parcijalne izvode po α_k i β_l i primeniti gradijentni spust na $-\log P(\mathbf{y}|\mathbf{x})$

$$\frac{\partial \log P}{\partial \alpha_k} = -\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \alpha_k}(\mathbf{y}-\boldsymbol{\mu}) + \left(\frac{\partial \mathbf{b}^T}{\partial \alpha_k} - \boldsymbol{\mu}^T \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \alpha_k}\right)(\mathbf{y}-\boldsymbol{\mu}) + \frac{1}{2} Tr\left(\boldsymbol{\Sigma} \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \alpha_k}\right)$$

$$\frac{\partial \log P}{\partial \beta_k} = -\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \beta_l}(\mathbf{y}-\boldsymbol{\mu}) + \frac{1}{2} Tr\left(\boldsymbol{\Sigma} \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \beta_l}\right)$$

Učenje

- Gradijentni spust neće garantovati da će α_k i β_l biti strogo pozitivni
- Moguće je optimizovati po novim promenljivim $u_k = \log \alpha_k$ i $v_k = \log \beta_l$
- Jedna iteracija gradijentnog spusta košta $O(N^3)$, gde je N veličina trening skupa, zbog inverzije matrice $\mathbf{\Sigma}^{-1}$
- Ukoliko je matrica $\mathbf{\Sigma}^{-1}$ retka, onda je cena jedne iteracije $O(N^2)$

Primene

- Predviđanje koncentracije aerosola u vazduhu

Table 1. RMSE and FRAC of C005, NN, and NN+CCRF using features defined over five regions, without ($\beta = 0$) and with spatio-temporal correlation ($\beta \neq 0$)

	RMSE				FRAC			
Region	C005	NN	CRF,$\beta = 0$	CRF,$\beta \neq 0$	C005	NN	CRF,$\beta = 0$	CRF,$\beta \neq 0$
Whole Globe	0.123	0.112±0.001	0.1067±0.0005	0.1056±0.0005	0.65	0.667±0.002	0.704±0.003	0.708±0.004
N. America	0.098	0.085±0.001	0.083±0.001	0.081±0.001	0.64	0.667±0.008	0.71±0.01	0.71±0.01
S. America	0.140	0.110±0.005	0.104±0.003	0.098±0.002	0.55	0.56±0.01	0.58±0.02	0.60±0.02
Europe	0.080	0.080±0.001	0.0736±0.0005	0.0728±0.0005	0.76	0.762±0.005	0.807±0.006	0.812±0.006
Africa	0.172	0.154±0.001	0.152±0.001	0.149±0.001	0.53	0.560±0.006	0.568±0.007	0.577±0.006
Asia&Aus.	0.161	0.156±0.001	0.145±0.001	0.148±0.001	0.64	0.66±0.01	0.71±0.01	0.70±0.01

Pregled

CCRF

GCRF

FF-GCRF

Slučaj gustih grafova zavisnosti

- Velika složenost učenja i predviđanja onemogućava primenu na gusto povezane grafove, odnosno probleme u kojima postoje zavisnosti među većinom promenljivih
- Šta raditi?

Slučaj gustih grafova zavisnosti

- Velika složenost učenja i predviđanja onemogućava primenu na gusto povezane grafove, odnosno probleme u kojima postoje zavisnosti među većinom promenljivih
- Šta raditi?
- Aproksimacijama sniziti računsku složenost

Teorija usrednjenog polja (eng. mean field theory)

- Problem interakcija velikog broja elemenata nekog sistema je po pravilu težak problem
- U fizici se kao rešenje koristi aproksimacija pri kojoj se uticaj svih pojedinačnih elemenata na jedan konkretan element zameni njihovim ukupnim uticajem (uprosečenim poljem) na taj element
- Kako tačno podesiti taj ukupni uticaj tako da rezultati budu dobri?
- Ovaj princip je više puta oktrivan i u drugim disciplinama

Aproksimacija usrednjenim poljem (eng. mean field approximation)

- Neka je:

$$Q(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{N} Q_i(y_i|\mathbf{x})$$

- Pod tom pretpostavkom, rešiti

$$\min_{Q} D_{KL}(Q||P)$$

- Rešenje:

$$\log(Q_i(y_i|\mathbf{x})) = E_{y_1 \dots y_{i-1}, y_{i+1} \dots y_N}[\log P(\mathbf{y}|\mathbf{x})] + const$$

Aproksimacija usrednjenim poljem (eng. mean field approximation)

- U slučaju GCRF-a, to daje:

$$\log(Q_i(y_i|\mathbf{x})) = -\sum_{k=1}^{K}\alpha_k(y_i^2 - 2y_i R_{ik}(\mathbf{x})) - 2\sum_{l=1}^{L}\beta_l \sum_{i\neq j} S_{ij}^l(\mathbf{x})(y_i^2 - 2y_i E[y_j]) + const$$

- Svaka $Q_i(y_i|\mathbf{x})$ je Gausova raspodela sa parametrima:

$$\mu_i = \frac{\sum_{k=1}^{K}\alpha_k R_{ik}(\mathbf{x}) + 2\sum_{l=1}^{L}\beta_l \sum_{j\neq i} S_{ij}^l(\mathbf{x})\mu_j}{\sum_{k=1}^{K}\alpha_k + 2\sum_{l=1}^{L}\beta_l \sum_{j\neq i} S_{ij}^l(\mathbf{x})}$$

$$\sigma_i^2 = \frac{1}{2\left(\sum_{k=1}^{K}\alpha_k + 2\sum_{l=1}^{L}\beta_l \sum_{j\neq i} S_{ij}^l(\mathbf{x})\right)}$$

- Složenost iterativnog određivanja μ_i je $O(IN^2)$ pri pretpostavci $I << N$ za broj iteracija
- Složenost računanja varijansi je $O(N^2)$

Gausov kernel

- Neka je

$$S_{ij}^l(\mathbf{x}) = k_l(p_i^l, p_j^l) = \exp\left(-\frac{1}{2}(p_i^l - p_j^l)^T \Lambda_l (p_i^l - p_j^l)\right)$$

  gde je p_i^l skup atributa instance i u nekom prostoru (koji može zavisiti od l), a Λ_l simetrična pozitivno definitna matrica

- Sume koje uključuju sličnost se sada mogu zapisati kao:

$$\sum_{j \neq i} S_{ij}^l(\mathbf{x})\mu_j = \sum_j k_l(p_i^l, p_j^l)\mu_j - \mu_i$$

$$\sum_{j \neq i} S_{ij}^l(\mathbf{x}) = \sum_j k_l(p_i^l, p_j^l)\mu_j - 1$$

Konvolucija sa Gausovim kenrelom

- Ključni posao je izračunati konvoluciju:

$$\sum_j k_l(p_i^l, p_j^l)\mu_i$$

koja se naivno računa u vremenu $O(N^2)$

- Postoje efikasne aproksimacije!

Brza konvolucija sa Gausovim kernelom

- Konstruisati regularnu rešetku (postoje različiti modeli) koja pokriva prostor nad kojim je definisan kernel
- Konvolucija na rešetci samo sa bliskim susedima se računa u vremenu $O(d^2N)$ gde je d dimenzija vektora p_i^l, ali samo ukoliko se radi sa kernelom čija je matrica Λ_l identička
- Šta činiti?

Brza konvolucija sa Gausovim kernelom

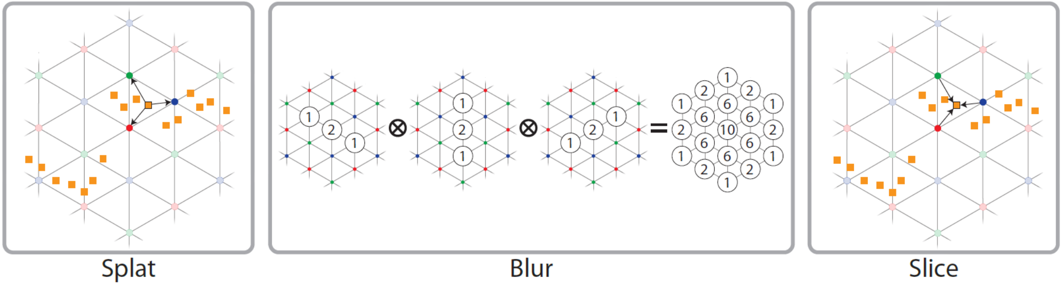
- Dekomponovati $\Lambda_l = U^T U$ Čoleski dekompozicijom (dimenzija Λ_l je mala u odnosu na broj promenljivih)

$$k_l(p_i^l, p_j^l) = \exp\left(-\frac{1}{2}(p_i^l - p_j^l)^T U^T I U (p_i^l - p_j^l)\right)$$

$$= \exp\left(-\frac{1}{2}(Up_i^l - Up_j^l)^T I (Up_i^l - Up_j^l)\right)$$

- Očito, primena polaznog kernela nad vektorima p_i^l je isto što i primena kernela sa identičkom matricom nad vektorima Up_i^l
- Transformisati sve p_i^l u Up_i^l i projektovati na temena rešetke
- Uraditi konvoluciju sa Gausovim kernelom pretpostavljajući nezavisnost duž pravaca rešetke, a uzimajući u obzir samo temena u nekoj uskoj okolini
- Aproksimirati vrednosti konvolucije u tačkama Up_i^l na osnovu vrednosti na rešetci

Brza konvolucija sa Gausovim kernelom



Predviđanje

- Predviđanje

$$y_i = \arg\max_y Q_i(y|\mathbf{x}) = \mu_i$$

- Korišćenjem brze konvolucije, složenost je $O(Id^2N)$
- I je fiksirano na primer na 200

Učenje

- Neka je:

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \log P(\mathbf{y}|\mathbf{x})$$

- Važi:

$$\frac{\partial \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \alpha_k} = \sum_{i=1}^{N} \left(E_Q[(y_i - R_{ik}(\mathbf{x}))^2] - (y_i - R_{ik}(\mathbf{x}))^2 \right)$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \beta_l} = \sum_{i=1}^{N} \sum_{i \neq j} k_l(p_i^l, p_j^l) \left(E_Q[(y_i - y_j)^2] - (y_i - y_j)^2 \right)$$

- Ispostavlja se da se brzom konvolucijom gradijent računa u vremenu $O((Id^2 + K + LId^2)N)$

Primene

- Predviđanje koncentracije aerosola u vazduhu

Table 1: Root mean squared error (RMSE) of fast fully connected CRF (FF-CCRF), fully-connected CCRF, and domain-knowledge based algorithm (C005) for aerosol optical depth prediction on different test set sizes N.

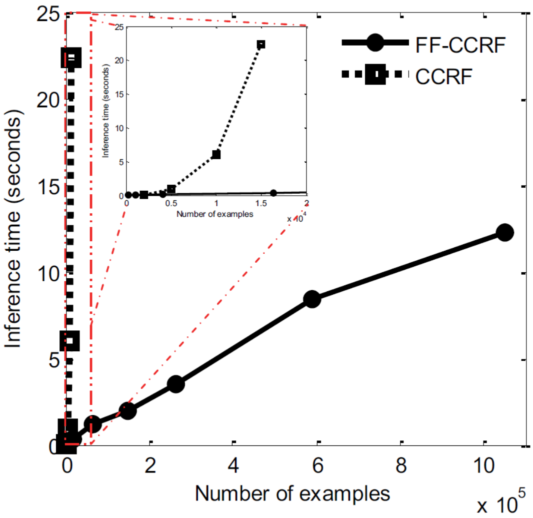
SIZE	$N = 3,000$	$N = 20,000$
FF-CCRF	$\mathbf{0.117 \pm 0.001}$	$\mathbf{0.115 \pm 0.005}$
C005	0.126 ± 0.010	0.129 ± 0.010
CCRF	$\mathbf{0.116 \pm 0.001}$	N/A

- Uklanjanje šuma sa slika

Table 2: Average root mean squared error (RMSE) over 68 test images as result of denoising by FF-CCRF and NML (lower RMSE is better).

NOISE LEVEL	10	15	20	25
NLM	6.20	8.11	9.39	10.68
FF-CCRF	**6.05**	**7.85**	**9.17**	**10.43**

Brzina



FF-CCRF
CCRF
Inference time (seconds)
Number of examples
x 10^5
0
2
4
6
8
10
25
20
15
10
5
Inference time (seconds)
Number of examples
x 10^4
0
0.5
1
1.5
2

Literatura

- V. Radosavljević, S. Vučetić, Z. Obradović, Continuous Conditional Random Fields for Regression in Remote Sensing
- V. Radosavljević, S. Vučetić, Z. Obradović, Neural Gaussian Conditional Random Fields
- K. Ristovski, V. Radosavljević, S. Vučetić, Z. Obradović, Continuous Conditional Random Fields for Efficient Regression in Large Fully Connected Graphs
- A. Adams, J. Baek, M. A. Davis, Fast High-Dimensional Filtering Using the Permutohedral Lattice
- P. Krähenbühl, V. Koltun, Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials
- J. Glass, M. Ghalwash, M. Vukićević, Z. Obradović, Extending the Modelling Capacity of Gaussian Conditional Random Fields while Learning Faster

HVALA NA PAŽNJI!