Few Shot Learning

Machine Learning that scales

Machine Learning and Applications Group, 2019.

Uroš Stegić

urosstegic@gmx.com

INTRODUCTION

General Overview

Examples

Basic Principles

Problems

- Amount of labeled data
- Retraining for new examples
- Classes unavailable during training
- Large number of classes

Big Dataset Absurd

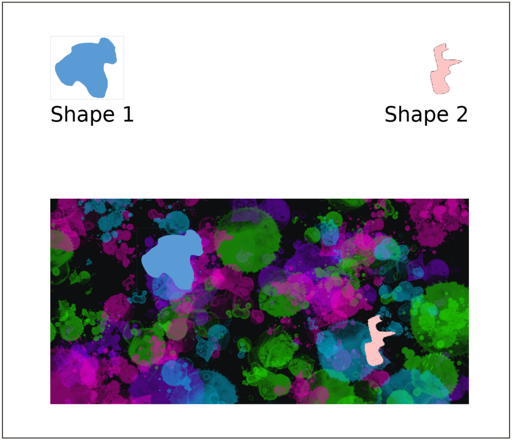

Figure: Shape recognition absurd

Person Re-Identification

Figure: Viewing the same scene with two cameras

Wake-up Word Detection

Figure: Learning Acoustic Word Embeddings [HYYH18]

Definition

One-shot *learning is an object categorization problem which aims to learn information about object categories from one, or only a few, training samples.*

Historical View

- Instance based algorithms (KNN)
- Mahalanobis distance [Mah36]
- Metrics learning
- Siamese neural networks [BGL+93]

K - Nearest Neighbours

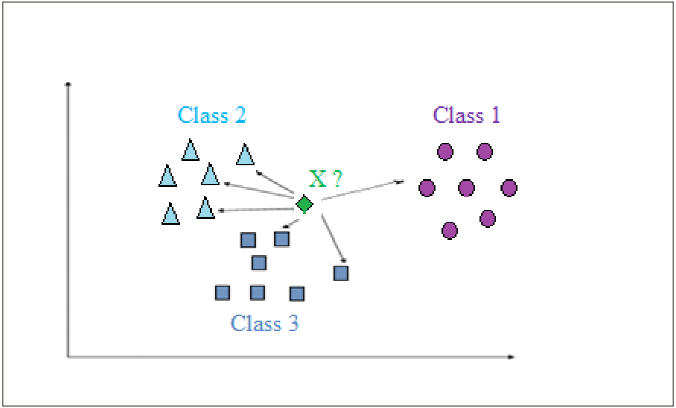

Figure: K - Nearest Neighbours

Mahalanobis Distance

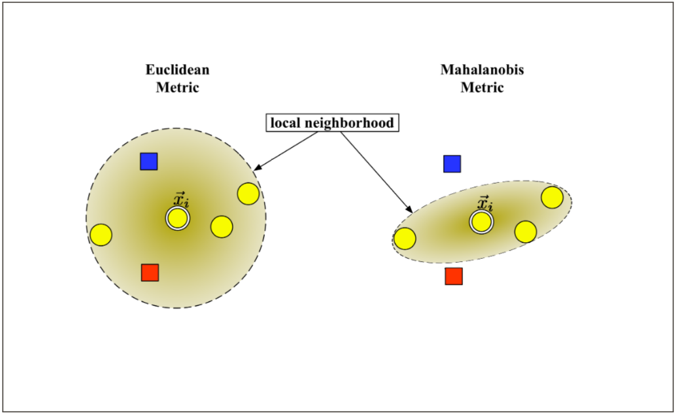

Figure: Local Neighbourhoods

Mahalanobis Distance - Formulation

Definition

Let $x, y \in \mathcal{D} \subseteq \mathbb{R}^n$ with the covariance matrix Σ. Mahalanobis distance $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined as follows:

$$d(x, y) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)}$$

Mahalanobis Distance - Properties

- Performs better
- Invariant to linear transformations
- Fast to compute
- Assumes linear correlation

Nonlinearity

*We're living in a **non-linear** world!*

Madonna

Embeddings

- Transform data in a non-linear manner
- Maintain semantic information
- Learn distance function [1]
- Measure distance on embedded objects

[1] or enforce embedding to be a euclidian space.

Siamese Architecture

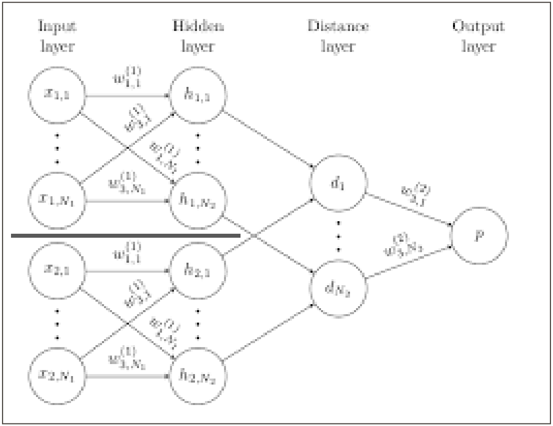

Figure: Siamese Neural Networks [GK15]

STANDARD TECHNIQUES

Contrastive Loss

Tripplet Loss

Quadruplet Loss

Matching Networks

More of Influental Papers

Metric Learning

- Probabilistic models vs. Energy-based models [LJH05]
- Low-dimensional embedding: $G_W(X)$
- Similarity metric: $E_W(X_1, X_2) = \|G_W(X_1) - G_W(X_2)\|$ [CHL05]

Siamese EBM

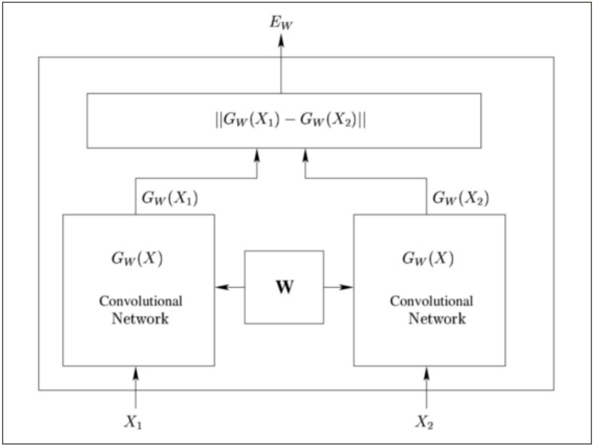

Figure: EBM model with siamese architecture

Contrastive Loss

- $\mathcal{L}(W) = \sum_{i=1}^{n} L(W, (Y, X_1, X2)^i)$

Contrastive Loss

- $\mathcal{L}(W) = \sum_{i=1}^{n} L(W, (Y, X_1, X2)^i)$
- $L(W, (Y, X_1, X2)^i) = (1 - Y)L_G(E_W(X_1, X_2)^i) + (Y)L_I(E_W(X_1, X_2)^i)$

Contrastive Loss

- $\mathcal{L}(W) = \sum_{i=1}^{n} L(W, (Y, X_1, X2)^i)$
- $L(W, (Y, X_1, X2)^i) = (1 - Y)L_G(E_W(X_1, X_2)^i) + (Y)L_I(E_W(X_1, X_2)^i)$
- $L_G(E_W(X_1, X_2)) = \frac{2}{Q}(E_W(X_1, X_2))^2$
- $L_I(E_W(X_1, X_2)) = 2Qe^{-\frac{2.77}{Q}E_W(X_1, X_2)}$

Tripplet Loss - Setup

Unified embedding for Face Recognition and Clustering [SKP15]

- Provide embedding
- Euclidian space
- Tripplets: anchor, positive, and negative

Tripplet Loss - Disclamer

Although we did not directly compare to other losses, we believe that the triplet loss is more suitable for face verification.

Tripplet Loss - Visualization

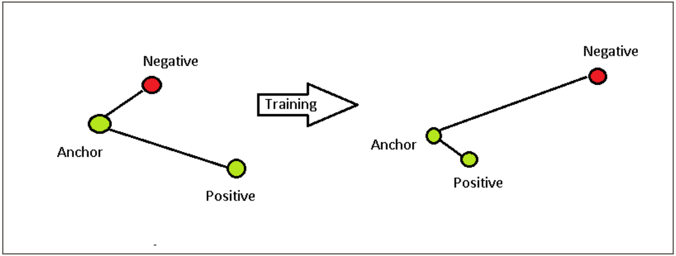

Figure: Tripplet loss - training objective

Tripplet Loss - Formulation

For i-th tripplet $(x_i^a, x_i^p, x_i^n) \in \mathcal{T}$ and embedding function f_W:

$$\left\| f_W(x_i^a) - f_W(x_i^p) \right\|_2^2 + \alpha < \left\| f_W(x_i^a) - f_W(x_i^n) \right\|_2^2$$

Tripplet Loss - Formulation

For i-th tripplet $(x_i^a, x_i^p, x_i^n) \in \mathcal{T}$ and embedding function f_W:

$$\mathcal{L}(W) = \sum_{i}^{|\mathcal{T}|} \left\| f_W(x_i^a) - f_W(x_i^p) \right\|_2^2 - \left\| f_W(x_i^a) - f_W(x_i^n) \right\|_2^2 + \alpha$$

Tripplet Mining

For given example x_i^a, the following instances are important:

- Hard positive: $x_i^{hp} = argmax_{x_i^p} \left\| f_W(x_i^a) - f_W(x_i^p) \right\|_2^2$
- Hard negative: $x_i^{hn} = argmin_{x_i^n} \left\| f_W(x_i^a) - f_W(x_i^n) \right\|_2^2$

Tripplet Mining

- Online mini-batch mining
- Fix the number of anchors per mini-batch (40)
- Include all positive examples in mini-batch
- Sample additional negative examples
- Use semi-hard negative examples (local minima)

Quadruplet Loss - Introduction

Beyond tripplet loss: a deep quadruplet network for person re-identification [CCZH17]

- Tripplets has low generalization
- QL gives smaller intra-class variation
- QL gives bigger inter-class variation
- Learn metric $g(x_i, x_j)$ instead of euclidian distance

QL - Embeddings

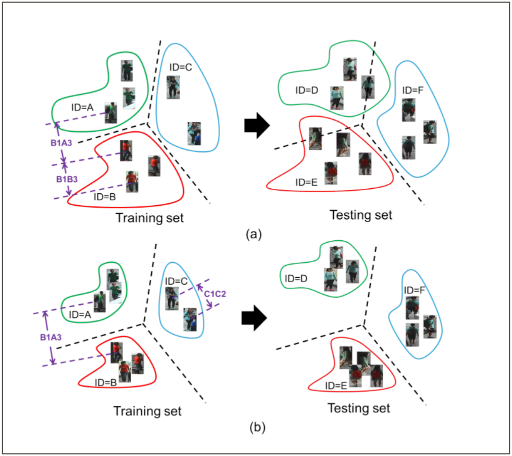

Figure: Embeddings with tripplet loss (a) vs. QL (b)

QL - Formulation

- Modified tripplet loss:

$$L_{trp} = \sum_{i,j,k}^{N} [g(x_i, x_j)^2 - g(x_i, x_k)^2 + \alpha_{trp}]_+$$

- Quadruplet loss:

$$L_{quad} = L_{trp} + \sum_{i,j,k,l}^{N} [g(x_i, x_j)^2 - g(x_l, x_k)^2 + \alpha_2]_+$$

Matching Networks - Introduction

- Using a support set: S
- Learn mapping: $S \rightarrow c(\hat{x})$
- Use attention mechanism over support set
- $\hat{y} = argmax_y P(y|\hat{x}, S)$

Matching Networks - Visualization

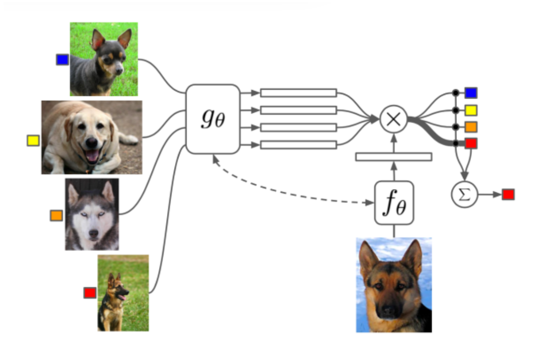

Figure 1: Matching Networks architecture

Figure: Matching Network Architecture

Matching Networks - Formulation

- $\hat{y} = \sum_{i=1}^{k} a(\hat{x}, x_i) y_i$
- $a(\hat{x}, x_i) = \frac{e^{c(f(\hat{x}), g(x_i))}}{\sum_{j=1}^{k} e^{c(f(\hat{x}), g(x_j))}}$

Matching Networks - Training

- Task T - distrubution over possible labels L
- Sample L from T
- Use L to sample S and B from the dataset
- Minimise error in batch B, given S

$$\theta = argmax_{\theta} E_{L \sim T}[E_{S \sim L, B \sim L}[\sum_{(x,y) \in B} \log P_{\theta}(y|x, S)]]$$

More of Influental Papers

- Lifted Structured Feature Embedding [SXJS15]
- Angular Loss [WZW+17]
- Cosine Metric Learning [WB18]
- In Defense of the Tripplet Loss [HBL17]

FIN

References I

- J. Bromley, I. Guyon, Y. Lecun, E. Sackinger, and R. Shah, *Signature verification using a siamese time delay neural network*, Advances in Neural Information Processing Systems (1993).
- Weihua Chen, Xiaotang Chen, Jianguo Zhang, and Kaiqi Huang, *Beyond triplet loss: A deep quadruplet network for person re-identification*.
- Sumit Chopra, Raia Hadsell, and Yann Lecun, *Learning a similarity metric discriminatively, with application to face verification*, Proc. Computer Vision and Pattern Recognition **1** (2005), 539– 546 vol. 1.

References II

- Ruslan Salakhutdinov Gregory Koch, Richard Zemel, *Siamese neural networks for one-shot image recognition*.
- Alexander Hermans, Lucas Beyer, and Bastian Leibe, *In defense of the triplet loss for person re-identification*, CoRR **abs/1703.07737** (2017).
- L. Hyungjun, K. Younggwan, J. Youngmoon, and K. Hoirin, *Learning acoustic word embeddings with phonetically associated triplet network*.
- Yann Lecun and Fu Jie Huang, *Loss functions for discriminative training of energy-based models*.

References III

- P.C. Mahalanobis, *On the generalized distance in statistics*, Proceedings of the National Institute of Science of India **2** (1936), 49–55.
- Florian Schroff, Dmitry Kalenichenko, and James Philbin, *Facenet: A unified embedding for face recognition and clustering*, CoRR **abs/1503.03832** (2015).
- Hyun Oh Song, Yu Xiang, Stefanie Jegelka, and Silvio Savarese, *Deep metric learning via lifted structured feature embedding*, CoRR **abs/1511.06452** (2015).
- Nicolai Wojke and Alex Bewley, *Deep cosine metric learning for person re-identification*, CoRR **abs/1812.00442** (2018).

References IV

Jian Wang, Feng Zhou, Shilei Wen, Xiao Liu, and Yuanqing Lin, *Deep metric learning with angular loss*, CoRR **abs/1708.01682** (2017).