Learning a metric for clustering structured data

Marek Dědič[1,2]

[1]ČVUT v Praze, Fakulta jaderná a fyzikálně inženýrská, Matematická informatika

[2]Cisco Systems Inc., Karlovo náměstí 10, Praha 2

October 11, 2019

Obsah

- Motivation
- MIL
- Contrastive predictive coding
- Triplet loss

Motivation

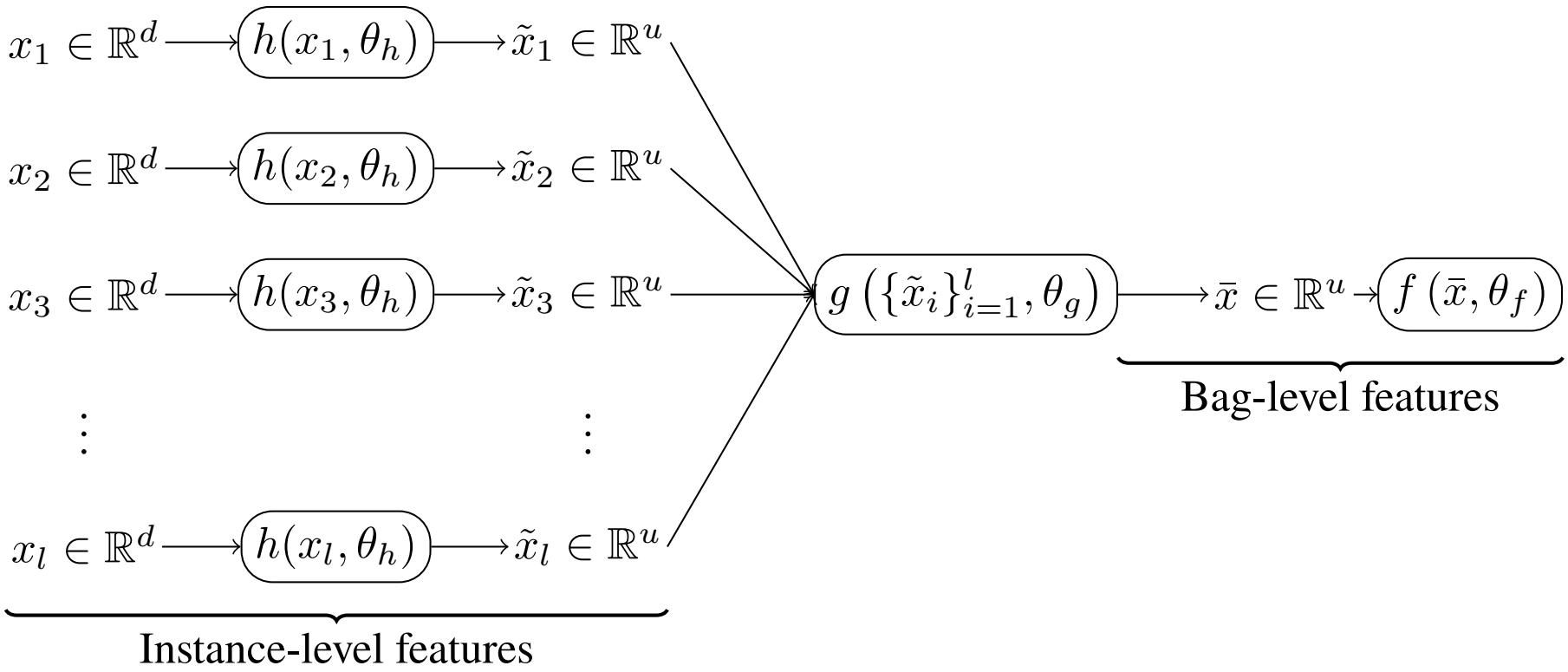
- Motivation
- MIL
- Contrastive predictive coding
- Triplet loss

Motivation

- Enable clustering of IPs from SwFlows - enhance Correlation
- Take advantage of the structure of the data using MIL
- Try out unsupervised learning

MIL

- Motivation
- MIL
- Contrastive predictive coding
- Triplet loss


$x_1 \in \mathbb{R}^d \longrightarrow h(x_1, \theta_h) \longrightarrow \tilde{x}_1 \in \mathbb{R}^u$
$x_2 \in \mathbb{R}^d \longrightarrow h(x_2, \theta_h) \longrightarrow \tilde{x}_2 \in \mathbb{R}^u$
$x_3 \in \mathbb{R}^d \longrightarrow h(x_3, \theta_h) \longrightarrow \tilde{x}_3 \in \mathbb{R}^u$
$x_l \in \mathbb{R}^d \longrightarrow h(x_l, \theta_h) \longrightarrow \tilde{x}_l \in \mathbb{R}^u$
Instance-level features
$g\left(\{\tilde{x}_i\}_{i=1}^{l}, \theta_g\right) \longrightarrow \bar{x} \in \mathbb{R}^u \rightarrow f\left(\bar{x}, \theta_f\right)$
Bag-level features

Contrastive predictive coding

- Motivation
- MIL
- Contrastive predictive coding
- Triplet loss

Contrastive predictive coding

The ideas of contrastive predictive coding were originally adapted giving a loss function

$$\log \left\| f\left(B_n^{(1)}\right) - f\left(B_n^{(2)}\right) \right\|^2 - \log \sum_{j=1}^{K} \left\| f\left(B_n^{(1)}\right) - f\left(B_j'\right) \right\|^2$$

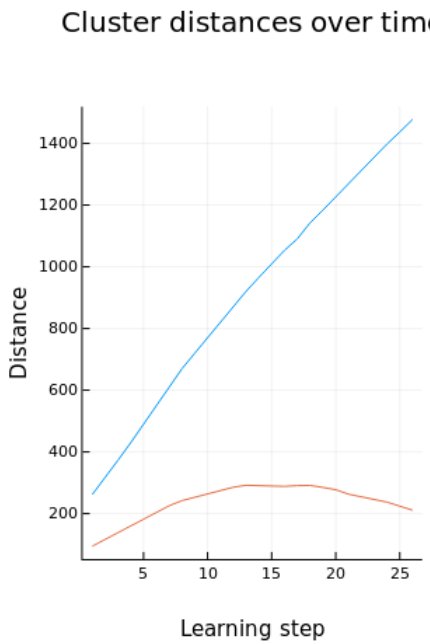
Contrastive predictive coding

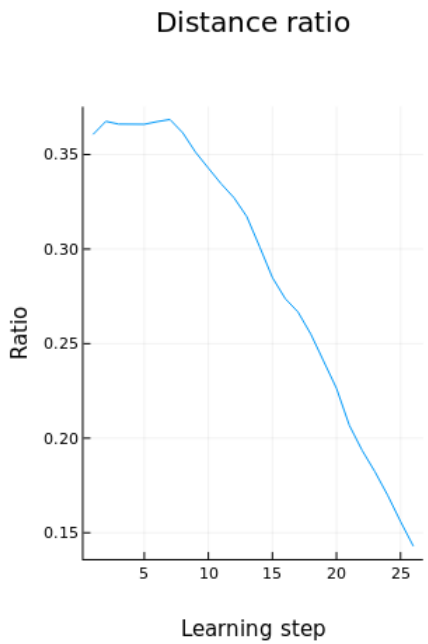
Later on, the loss function was simplified to

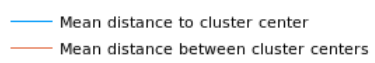
$$D_{ij} = \left\| f\left(B_i^{(1)}\right) - f\left(B_j^{(2)}\right) \right\|_2^2$$

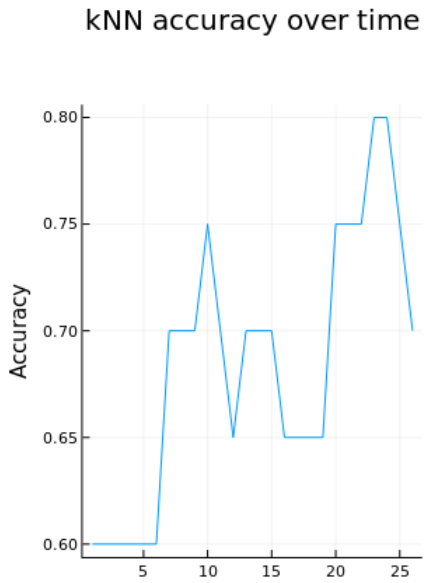
$$\frac{1}{n} \sum_{i=1}^{n} \left(\log\left(D_{ii}\right) - \log\left(\sum_{i \neq j} D_{ij} \right) \right)$$

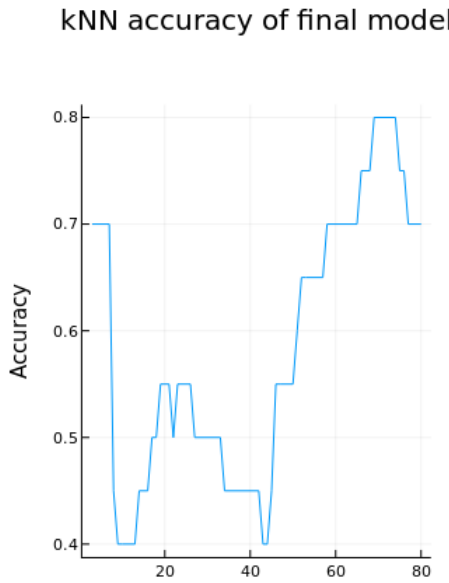
Contrastive predictive coding - preliminary results (Musk)











Triplet loss

- Motivation
- MIL
- Contrastive predictive coding
- Triplet loss

Triplet loss

The triplet loss is another alternative, this time requiring supervised learning

$$y_{ij} = \begin{cases} 1 & \text{for} \quad y_i = y_j \\ 0 & \text{otherwise} \end{cases}$$

$$\sum_{ij} y_{ij} D_{ij} + c \sum_{ijl} y_{ij} \left(1 - y_{il}\right) \max\left(0, 1 + D_{ij} - D_{il}\right)$$

Triplet loss - preliminary results (Musk)