Metric Learning

Why metric learning

Similarity

Unsupervised Learning

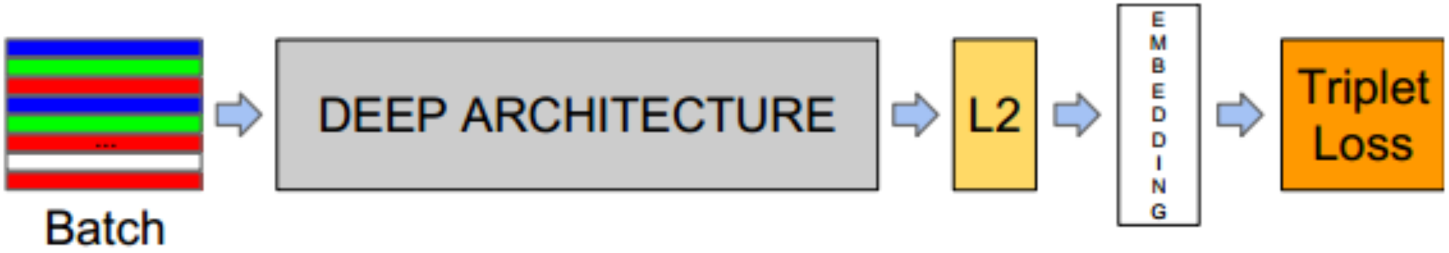
More Classes

Learning Distance Metrics

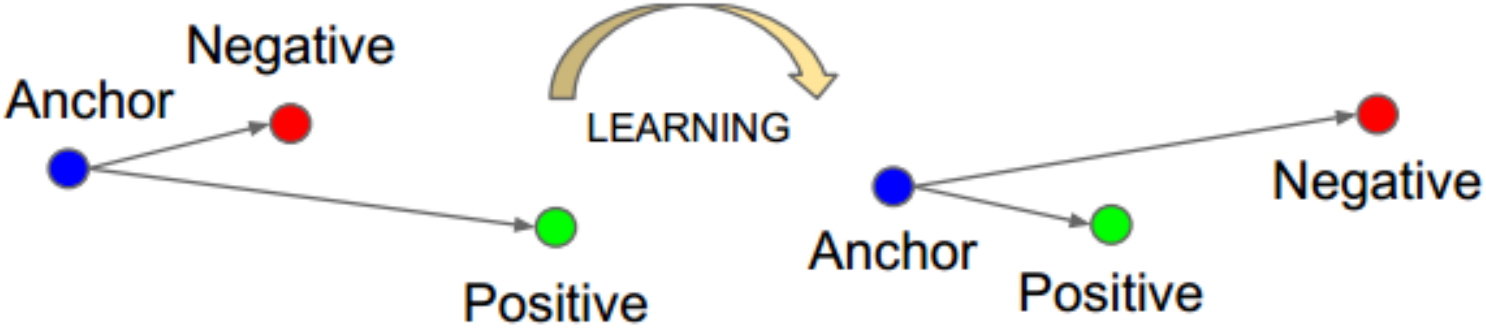
$$d(x, y) = d_A(x, y) = ||x - y||_A = \sqrt{(x - y)^T A (x - y)}.$$

$$\begin{array}{rl} \min\limits_{A} & \sum_{(x_i, x_j) \in \mathcal{S}} ||x_i - x_j||_A^2 \\ \text{s.t.} & \sum_{(x_i, x_j) \in \mathcal{D}} ||x_i - x_j||_A \geq 1, \\ & A \succeq 0. \end{array}$$

FaceNet



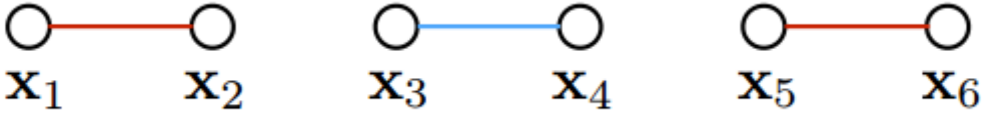
Triplet Loss



$$\|x_i^a - x_i^p\|_2^2 + \alpha < \|x_i^a - x_i^n\|_2^2, \ \forall \left(x_i^a, x_i^p, x_i^n\right) \in \mathcal{T}\ .$$

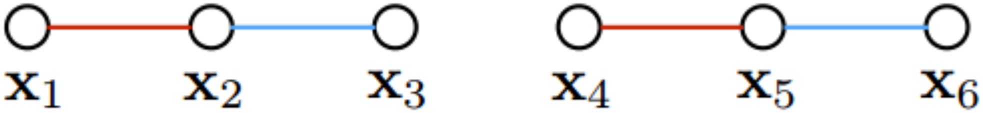
$$\sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

Lifted Structured Feature Embedding



(a) Contrastive embedding

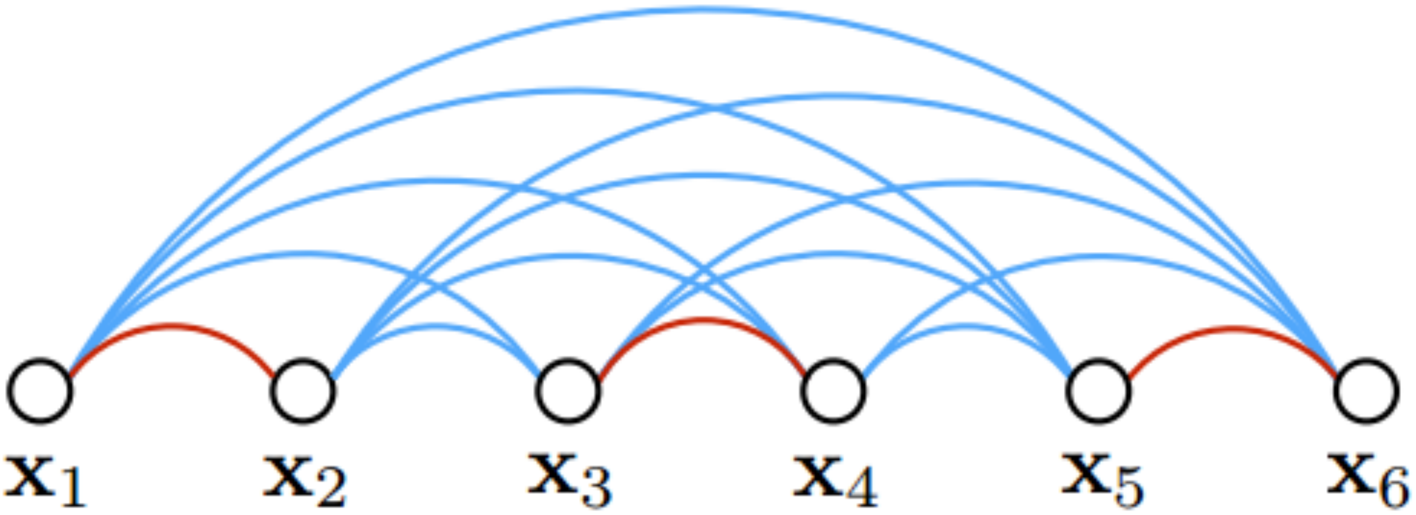
$$J = \frac{1}{m}\sum_{(i,j)}^{m/2} y_{i,j} D_{i,j}^2 + (1-y_{i,j})\left[\alpha - D_{i,j}\right]_+^2$$



(b) Triplet embedding

$$J = \frac{3}{2m}\sum_{i}^{m/3}\left[D_{ia,ip}^2 - D_{ia,in}^2 + \alpha\right]_+$$

Lifted Structured Feature Embedding



(c) Lifted structured embedding

$$J = \frac{1}{2|\widehat{\mathcal{P}}|} \sum_{(i,j)\in\widehat{\mathcal{P}}} \max\left(0,\ J_{i,j}\right)^2,$$

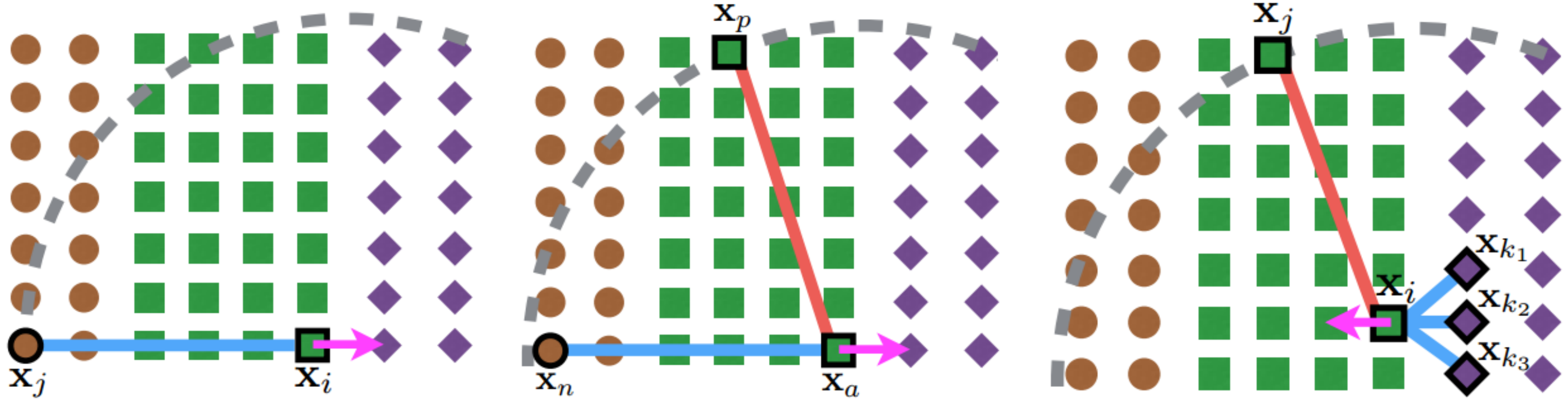
$$J_{i,j} = \max\left(\max_{(i,k)\in\widehat{\mathcal{N}}} \alpha - D_{i,k},\ \max_{(j,l)\in\widehat{\mathcal{N}}} \alpha - D_{j,l}\right) + D_{i,j}$$

Lifted Structured Feature Embedding

$$
\begin{aligned}
J &= \frac{1}{2|\widehat{\mathcal{P}}|} \sum_{(i,j)\in\widehat{\mathcal{P}}} \max\left(0,\ J_{i,j}\right)^2 , \\
J_{i,j} &= \max\left(\max_{(i,k)\in\widehat{\mathcal{N}}} \alpha - D_{i,k},\ \max_{(j,l)\in\widehat{\mathcal{N}}} \alpha - D_{j,l} \right) + D_{i,j}
\end{aligned}
$$

$$
\begin{aligned}
\tilde{J}_{i,j} &= \log\left(\sum_{(i,k)\in\mathcal{N}} \exp\{\alpha - D_{i,k}\} + \sum_{(j,l)\in\mathcal{N}} \exp\{\alpha - D_{j,l}\} \right) + D_{i,j} \\
\tilde{J} &= \frac{1}{2|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} \max\left(0,\ \tilde{J}_{i,j}\right)^2 , \qquad (4)
\end{aligned}
$$

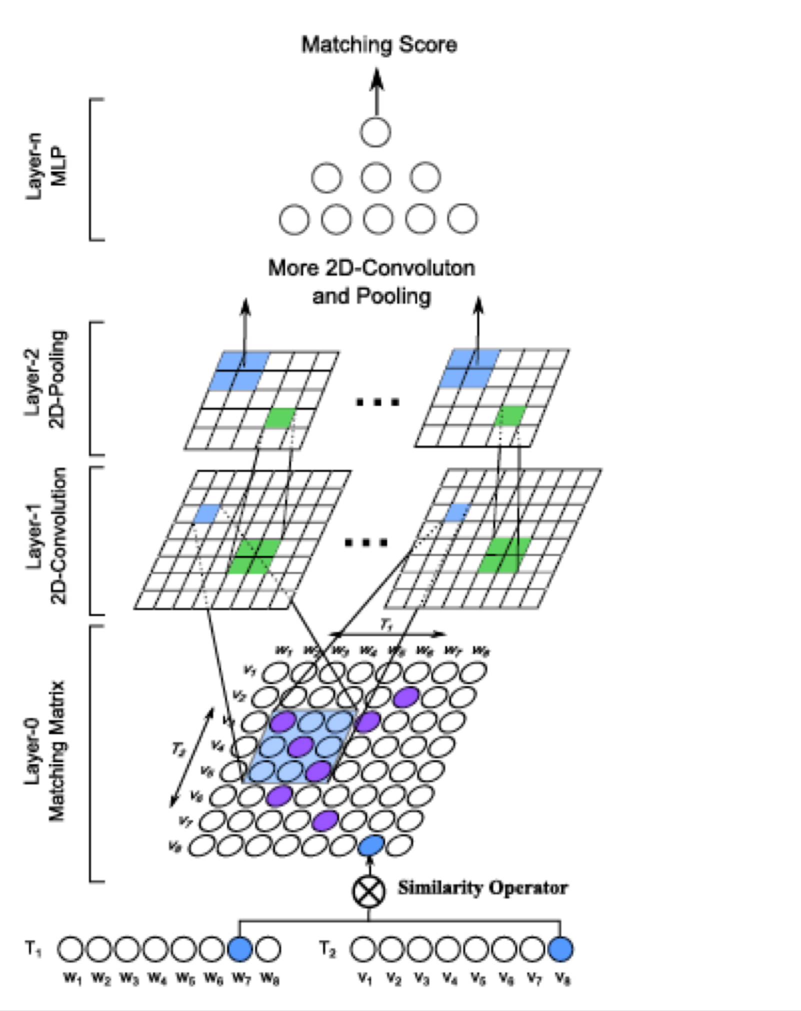
Lifted Structured Feature Embedding



(a) Contrastive embedding (b) Triplet embedding (c) Lifted structured similarity

Text Matching as Image Recognition

$$\mathrm{match}(T_1, T_2) = \mathrm{F}\big(\Phi(T_1), \Phi(T_2)\big)$$



Text Matching as Image Recognition

$$\mathbf{M}_{ij} = \mathbb{I}_{\{w_i = v_j\}} = \begin{cases} 1, & \text{if } w_i = v_j \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathbf{M}_{ij} = \frac{\vec{\alpha_i}^{\top} \vec{\beta_j}}{\|\vec{\alpha_i}\| \cdot \|\vec{\beta_j}\|},$$

$$\mathbf{M}_{ij} = \vec{\alpha_i}^{\top} \vec{\beta_j}.$$

Summary

Thanks

References:

[1] Distance Metric Learning, with Application to Clustering with Side-Information

[2] FaceNet: A Unified Embedding for Face Recognition and Clustering

[3] Deep Metric Learning via Lifted Structured Feature Embedding

[4] Text Matching as Image Recognition