Metric Learning Approaches for Face Identification

Computer Vision (CSE578) Course Project

Akshay Bankar (2019201011)

Angad Yennam (2019900082)

Face Identification as Metric Learning

- Metric learning : Constructing task-specific distance metrics from supervised data. The learned distance metric can then be used to perform task of clustering and classification.
- Mahalanobis distance : This can be thought as euclidean distance after a linear transformation of the feature space defined by L,

$$D_{\mathbf{L}}(\vec{x}_i, \vec{x}_j) = \|\mathbf{L}(\vec{x}_i - \vec{x}_j)\|_2^2.$$

- Mahalanobis metrics : Any matrix M formed from a real-valued matrix L such that,

$$\mathbf{M} = \mathbf{L}^\top \mathbf{L}.$$

Such a matrix formed is positive semidefinite.

- The squared distances then is denoted by,

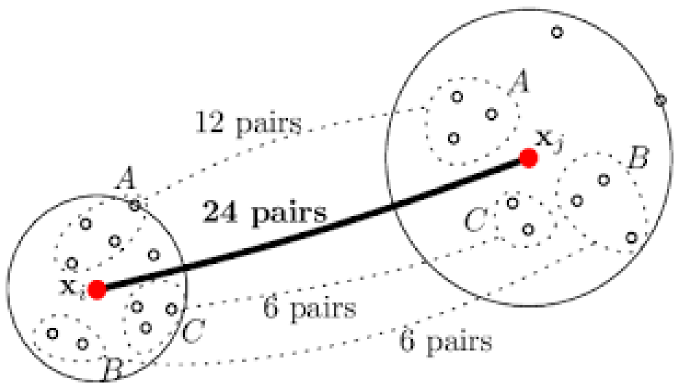
$$\mathcal{D}_{\mathbf{M}}(\vec{x}_i, \vec{x}_j) = (\vec{x}_i - \vec{x}_j)^\top \mathbf{M}(\vec{x}_i - \vec{x}_j).$$

Marginalized K-Nearest Neighbors (M-kNN) :

- To predict whether a pair of images (xi, xj) belongs to the same class or not marginal probability is computed that assigns xi and xj to the same class using a kNN classifier, which is given by

$$p(y_i = y_j | \mathbf{x}_i, \mathbf{x}_j) = \sum_c p(y_i = c|\mathbf{x}_i) p(y_j = c|\mathbf{x}_j)$$
$$= k^{-2} \sum_c n_c^i n_c^j.$$

- The probability for the positive class given by this classifier for a pair is determined by the number of positive and negative neighbour pairs.
- Hence the score of MkNN binary classifier for a pair of images (xi, xj) is based on how many positive neighbour pairs can be formed from neighbours ofxi and xj.

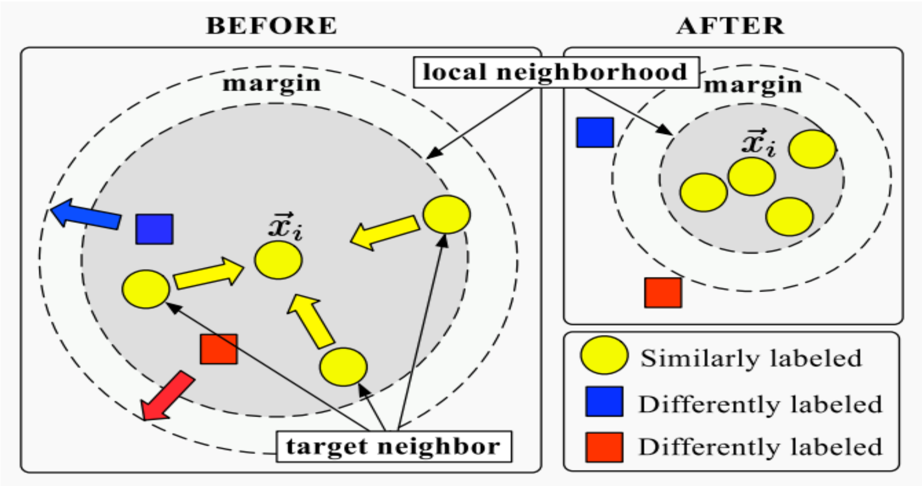


The author uses following methods to compare his methods :

1. <u>Large Margin Nearest Neighbors (LMNN)</u> :

- A distance metric learning algorithm for nearest neighbors' classification.
- It learns a metric that pulls the neighbor candidates (target_neighbors) near, while pushes near data from different classes (impostors) out of the target neighbors' margin.
- Target neighbors: Each input xi has k nearest neighbors that share its same label yi. These establisha perimeter based on Mahalanobis distance.
- Imposters : These are differently labeled inputs in the training set that invade perimeter set by thetarget neighbors.

$$\|\mathbf{L}(\vec{x}_i-\vec{x}_l)\|^2 \leq \|\mathbf{L}(\vec{x}_i-\vec{x}_j)\|^2+1.$$



Formulation :

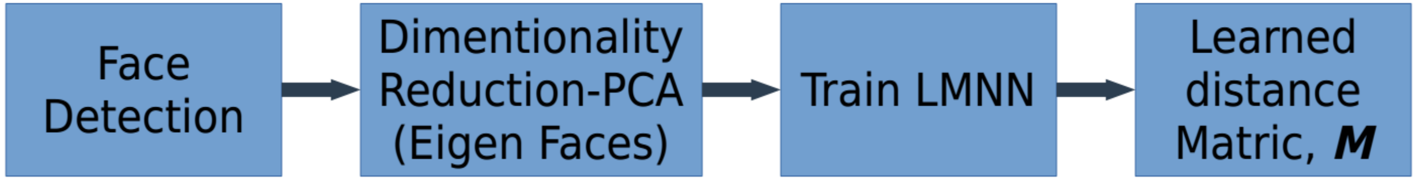
Minimize $(1-\mu)\sum_{i,j\rightsquigarrow i}(\vec{x}_i-\vec{x}_j)^\top \mathbf{M}(\vec{x}_i-\vec{x}_j)+\mu\sum_{i,j\rightsquigarrow i,l}(1-y_{il})\xi_{ijl}$ **subject to:**

(1) $(\vec{x}_i-\vec{x}_l)^\top \mathbf{M}(\vec{x}_i-\vec{x}_l)-(\vec{x}_i-\vec{x}_j)^\top \mathbf{M}(\vec{x}_i-\vec{x}_j) \ \geq\ 1-\xi_{ijl}$

(2) $\xi_{ijl}\geq 0$

(3) $\mathbf{M}\succeq 0$.

Implementation :



2. <u>Information Theoretic Metric Learning :</u>

- Given an initial metric, it learns the nearest metric that satisfies some similarity and dissimilarity constraints.
- The closeness between the metrics is measured using the Kullback-Leibler divergence between the corresponding gaussians.
- Given pairs of similar points S and pairs of dissimilar points D, it formulates distance metric learningproblem as,

$$\begin{array}{lll} \min\limits_{A} & \mathrm{KL}(p(\boldsymbol{x}; A_0) \| p(\boldsymbol{x}; A)) & \\ \text{subject to} & d_A(\boldsymbol{x}_i, \boldsymbol{x}_j) \leq u & (i,j) \in S, \\ & d_A(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq \ell & (i,j) \in D. \end{array}$$

$$\mathrm{KL}(p(\boldsymbol{x}; A_o) \| p(\boldsymbol{x}; A)) = \int p(\boldsymbol{x}; A_0) \log \frac{p(\boldsymbol{x}; A_0)}{p(\boldsymbol{x}; A)} \, d\boldsymbol{x}.$$

Formulation :

- LogDet divergence : The LogDet divergence is a Bregman matrix divergence generated by the convex function F(X) = log detX defined over the cone of positive-definite matrices.

$$D_{\ell d}(A, A_0) = \mathrm{tr}(AA_0^{-1}) - \log\det(AA_0^{-1}) - n.$$

- The differential relative entropy between two multivariate Gaussians can be expressed as the convex combination of a Mahalanobis distance between mean vectors and the LogDet divergence between the covariance matrices

$$\mathrm{KL}(p(\boldsymbol{x}; A_0)\|p(\boldsymbol{x}; A)) = \frac{1}{2}D_{\ell d}(A, A_0)$$

- Hence the optimization problem is re-posed as

$$\begin{aligned}
\min_{A\succeq 0,\boldsymbol{\xi}} \quad & D_{\ell d}(A, A_0) + \gamma \cdot D_{\ell d}(\mathrm{diag}(\boldsymbol{\xi}), \mathrm{diag}(\boldsymbol{\xi}_0)) \\
\text{s. t.} \quad & \mathrm{tr}(A(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T) \leq \xi_{c(i,j)} \quad (i,j) \in S, \\
& \mathrm{tr}(A(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T) \geq \xi_{c(i,j)} \quad (i,j) \in D,
\end{aligned}$$

Logistic Discriminant based Metric Learning (LDML) :

- A distance metric learning algorithm that maximizes the likelihood of a logistic based probability distribution.
- The distance between images in positive pairs to be smaller than the distances corresponding to negative pairs and obtain a probabilistic estimation of whether the two images depict the same object.
- The method models the probability that pair n = (i, j) is positive,

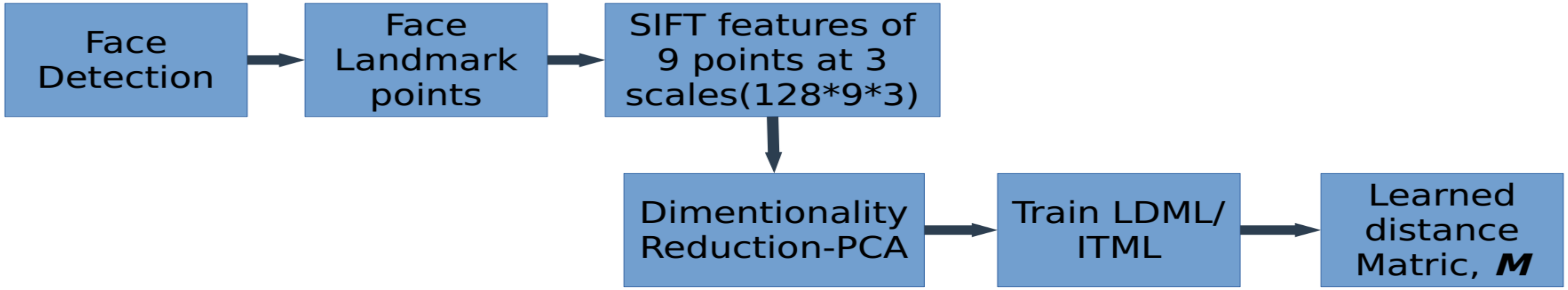
$$p_n = p(y_i = y_j | \mathrm{x}_i, \mathrm{x}_j; \mathrm{M}, b) = \sigma(b - d_{\mathrm{M}}(\mathrm{x}_i, \mathrm{x}_j))$$

- where $\sigma(z) = (1 + \exp(-z))^{-1}$

- To optimize the parameters, it uses maximum log-likelihood of L which can be written as :

$$\mathcal{L} = \sum_n t_n \ln p_n + (1 - t_n) \ln(1 - p_n)$$
$$\nabla \mathcal{L} = \sum_n (t_n - p_n) X_n,$$

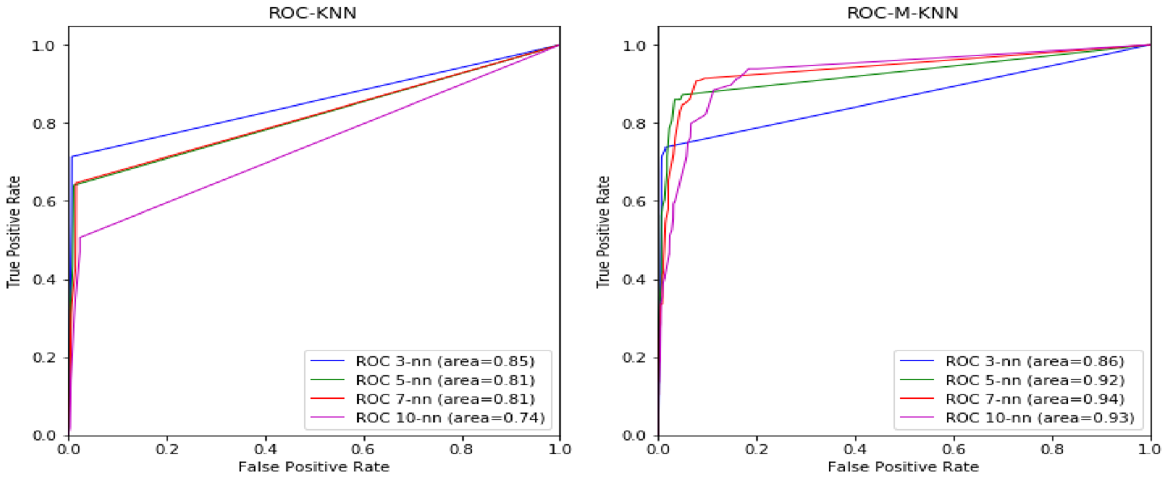
- This log-likelihood is smooth and concave. Thus, can be solved by gradient ascent.



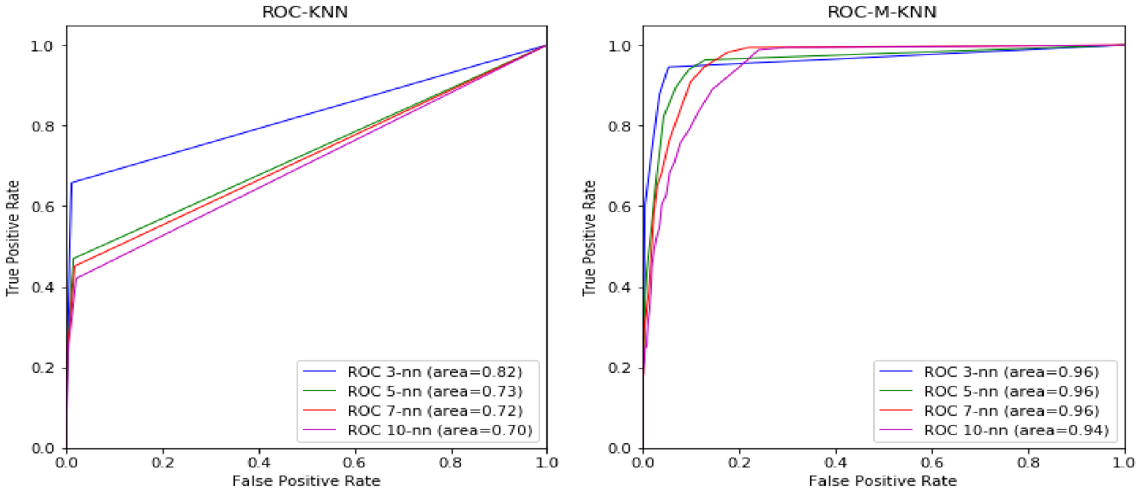
Results and Observations:

- Olivetti faces dataset :

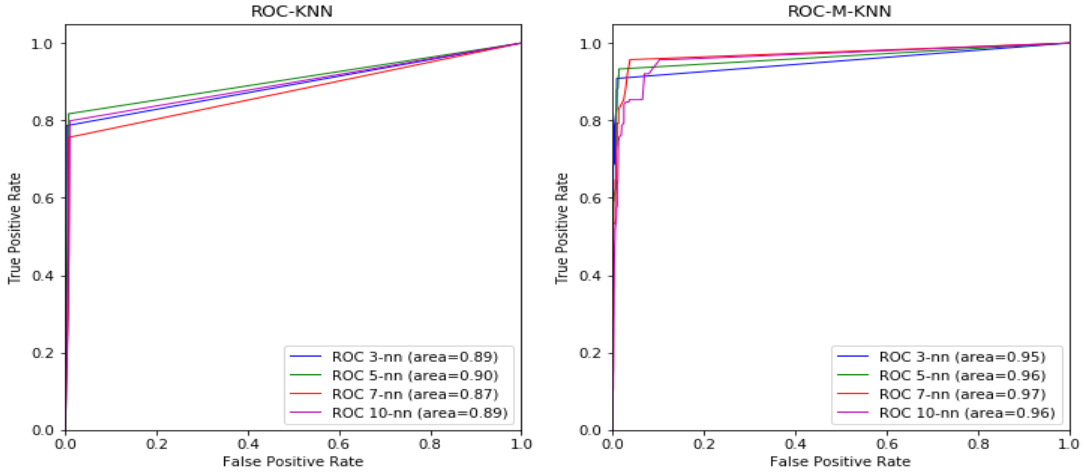
Method	NN-method	ROC classification results			
		3-NN	5-NN	7-NN	10-NN
LMNN	K-NN	85.28	81.45	81.47	74.14
	M-KNN	86.38	92.47	94.69	92.94
ITML	K-NN	89.11	90.41	87.23	89.28
	M-KNN	95.20	96.31	97.14	96.20
LDML	K-NN	82.39	72.85	71.83	70.15
	M-KNN	96.20	95.85	96.21	94.43



LMNN



LDML



ITML

Observations :

- The results show that the proposed method of LDML out-performs the LMNN and ITML methods used for comparison.

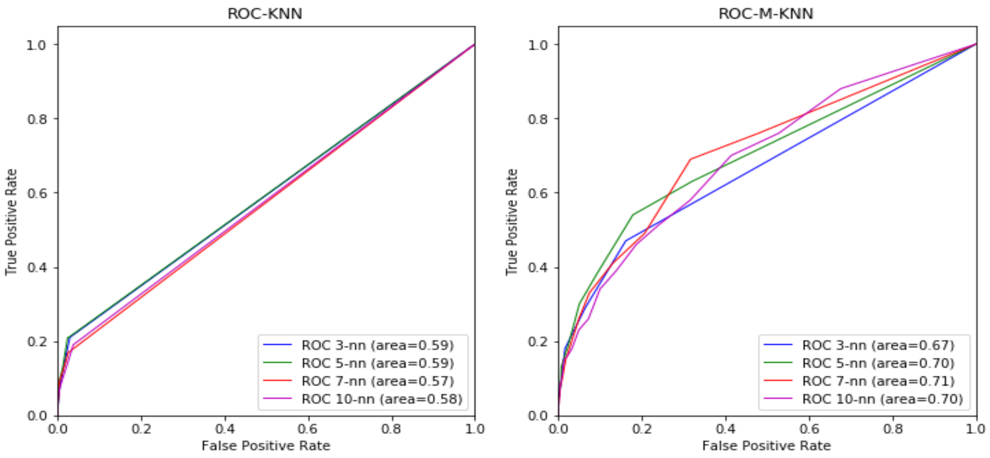
- The classification using proposed M-KNN improves results of all the methods as compared to when used with KNN.

- The improvement in result can be described by the fact that the dataset is less challenging in terms of variations in various imaging aspects.

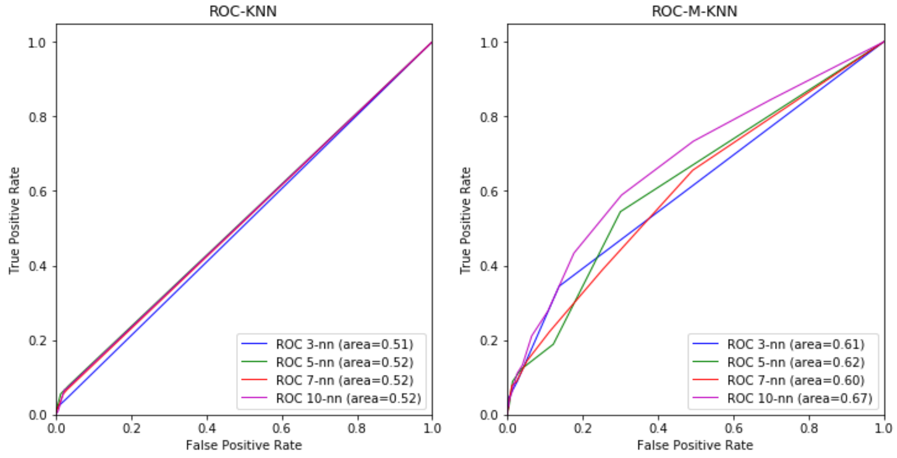
- But the proposed method gives comparatively sub-optimal results on more challenging LFW dataset.

- Labelled Faces in Wild dataset :

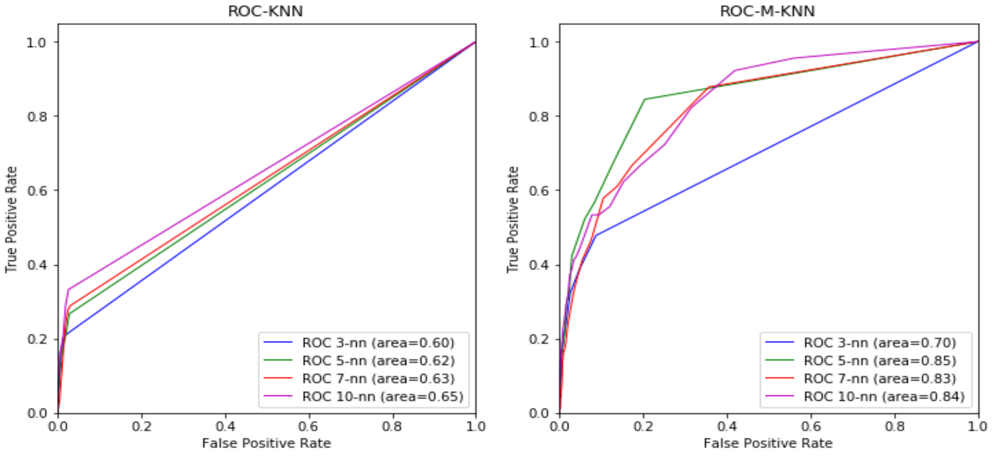
Method	NN-method	ROC classification results			
		3-NN	5-NN	7-NN	10-NN
LMNN	K-NN	59.01	59.23	57.23	57.70
	M-KNN	66.63	70.07	71.10	69.55
ITML	K-NN	59.62	62.03	62.98	65.41
	M-KNN	70.48	85.38	82.65	84.10
LDML	K-NN	50.76	52.25	51.90	52.09
	M-KNN	60.05	62.21	60.43	67.29



LMNN



LDML



ITML

Observations :

- The performance of LDML method on this dataset are not good w.r.t other two methods.

- The dataset, in general, is challenging, having a big variety in pose, expression, lighting as compared to Olivetti faces dataset.

- The M-KNN method shows consistent improvement in all methods' results when compared with K-NN as classification technique.

- This can be regarded to the fact that the score of Marginalized kNN (M-KNN) binary classifier for a pair of images (xi, xj) is based on how many positive neighbor pairs can be formed from neighbors of xi and xj. It is not "local" in the sense of usual K-NN classifiers as M-KNN measures the correspondence between two distinct local neighborhoods.

THANK YOU