Enhanced Eigenspace Separation Transform for Classification

K. Ramachandra Murthy

Indian Statistical Institute,
Kolkata

Outline

Outline

Outline

Outline

Objective

Supervised Feature Extraction

Reduction of dimensionality while preserving as much of the class discriminatory information as possible

Fisher Linear Discriminant Analysis

- Fisher suggested maximizing the difference between the means, normalized by a measure of the within-class scatter
- The classical Fisher linear discriminant is a transformation matrix that is the projection vector

$$U = S_w^{-1}(\mu_2 - \mu_1)$$

where μ_1 and μ_2 are the mean vectors of the two classes defined by

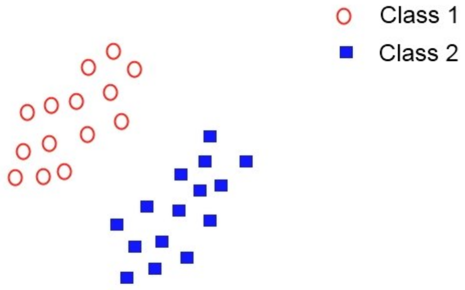
$$\mu_i = \frac{1}{m_i}\sum_{j=1}^{m_i} x_{ij}, \ \ i = 1,2$$

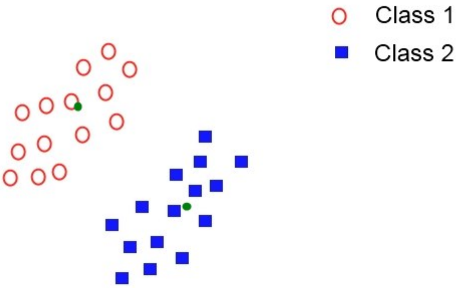
and S_w is the within-class covariance matrix defined by

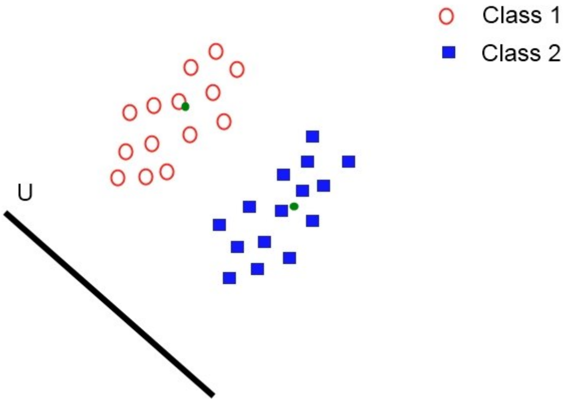
$$S_w = \frac{1}{m_1}\sum_{i=1}^{m_1}(x_{1i} - \mu_1)(x_{1i} - \mu_1)^T + \frac{1}{m_2}\sum_{i=1}^{m_2}(x_{2i} - \mu_2)(x_{2i} - \mu_2)^T$$

where m_1 and m_2 are the number of patterns from class1 and class2.

- Of all the possible lines we would like to select the one that maximizes the separability of the scalars


Class 1
Class 2


Class 1
Class 2



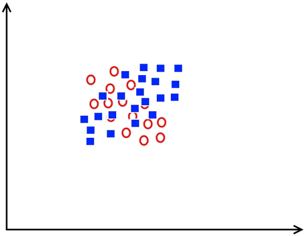
Note: Fisher Linear Discriminant Analysis can be extended to multi-class.

Disadvantages

The class means are equal.

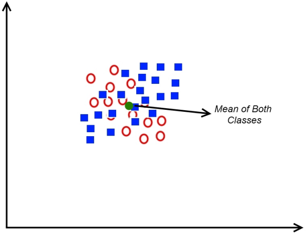
Disadvantages

The class means are equal.



Disadvantages

The class means are equal.



Disadvantages

The class means are equal.

The matrix S_w is singular or ill conditioned.

Disadvantages

The class means are equal.

The matrix S_w is singular or ill conditioned.

The data vectors are of high dimensionality and the projection onto a single dimension removes too much information.

Disadvantages

The class means are equal.

The matrix S_w is singular or ill conditioned.

The data vectors are of high dimensionality and the projection onto a single dimension removes too much information.

Note: Huge number of methods are existed in the literature to overcome the singularity problem.

Goal

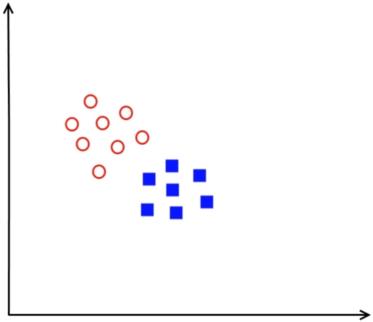
Form a suitable performance criterion for a transform that supports a binary classifier that maximizes some measure of the subspace separation of the two classes of data vectors.

Eigenspace Separation Transform (EST)

To separate the two classes of data vectors, EST seeks a subspace projection that increases the difference in the average lengths of the vectors in the two classes.

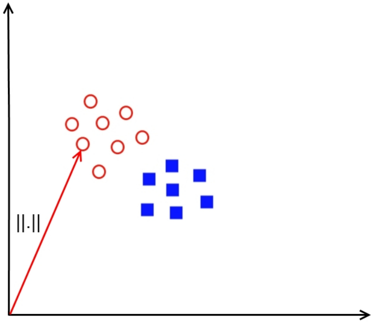
Don Torrieri, The eigenspace separation transform for neural network classifiers, Neural Networks, Volume 12, Issue 3, April 1999, Pages 419-427.

Eigenspace Separation Transform (EST)



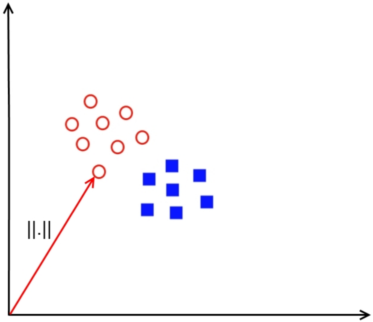
Don Torrieri, The eigenspace separation transform for neural network classifiers, Neural Networks, Volume 12, Issue 3, April 1999, Pages 419-427.

Eigenspace Separation Transform (EST)



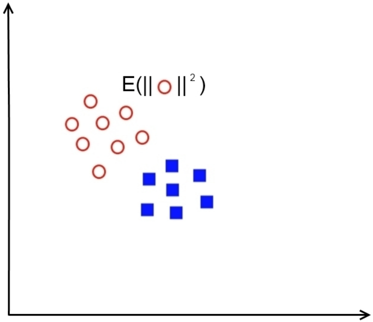
Don Torrieri, The eigenspace separation transform for neural network classifiers, Neural Networks, Volume 12, Issue 3, April 1999, Pages 419-427.

Eigenspace Separation Transform (EST)



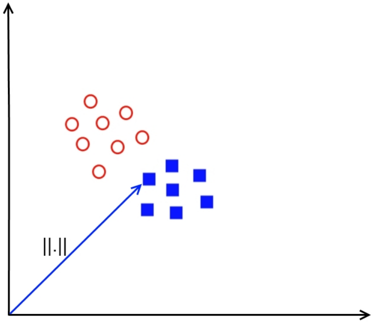
Don Torrieri, The eigenspace separation transform for neural network classifiers, Neural Networks, Volume 12, Issue 3, April 1999, Pages 419-427.

Eigenspace Separation Transform (EST)



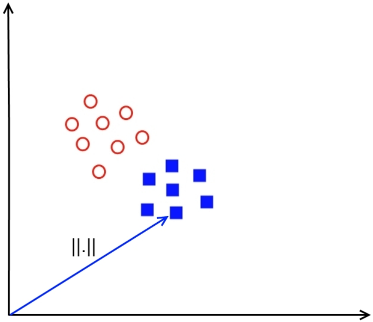
Don Torrieri, The eigenspace separation transform for neural network classifiers, Neural Networks, Volume 12, Issue 3, April 1999, Pages 419-427.

Eigenspace Separation Transform (EST)



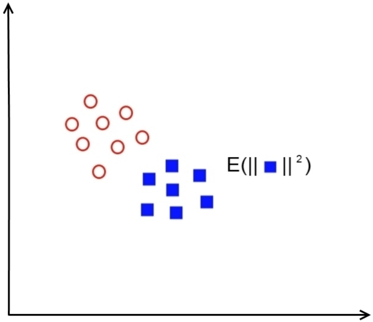
Don Torrieri, The eigenspace separation transform for neural network classifiers, Neural Networks, Volume 12, Issue 3, April 1999, Pages 419-427.

Eigenspace Separation Transform (EST)



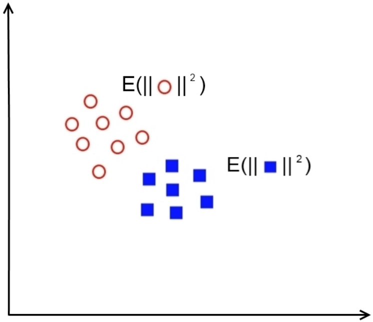
Don Torrieri, The eigenspace separation transform for neural network classifiers, Neural Networks, Volume 12, Issue 3, April 1999, Pages 419-427.

Eigenspace Separation Transform (EST)



Don Torrieri, The eigenspace separation transform for neural network classifiers, Neural Networks, Volume 12, Issue 3, April 1999, Pages 419-427.

Eigenspace Separation Transform (EST)



Don Torrieri, The eigenspace separation transform for neural network classifiers, Neural Networks, Volume 12, Issue 3, April 1999, Pages 419-427.

Objective

$$\max \left| E\left[\|\vec{y}_1\|^2 - \|\vec{y}_2\|^2 \right] \right|$$

Where $\vec{y}_i = W^T \vec{x}_i$, $i = 1, 2$; U is the transformation matrix and $\vec{x}_i$ denotes n-dimensional random vector of class i.

Objective

$$\max \left| E\left[\|\vec{y}_1\|^2 - \|\vec{y}_2\|^2 \right] \right|$$

Where $\vec{y}_i = W^T \vec{x}_i$, $i = 1, 2$; U is the transformation matrix and $\vec{x}_i$ denotes n-dimensional random vector of class i.

$$\text{Let } \; D = \left| E\left[\|\vec{y}_1\|^2 - \|\vec{y}_2\|^2 \right] \right|$$

Objective

$$\max \left| E\left[\|\vec{y}_1\|^2 - \|\vec{y}_2\|^2 \right] \right|$$

Where $\vec{y}_i = W^T \vec{x}_i$, $i = 1, 2$; U is the transformation matrix and $\vec{x}_i$ denotes n-dimensional random vector of class i.

$$\text{Let } \; D = \left| E\left[\|\vec{y}_1\|^2 - \|\vec{y}_2\|^2 \right] \right|$$

Substituting $\vec{y}_i = W^T \vec{x}_i$, $i = 1, 2$ in the above equation, we will get

$$D = \left| \sum_{i=1}^{k} \vec{w}_i^T M \vec{w}_i \right|.$$

Where $M = R_{x1} - R_{x2}$ and $R_{xi} = E\left[\vec{x}_i \vec{x}_i^T \right] \in \mathbb{R}^{n \times n}$, $\; i = 1, 2.$

Objective

$$\max \left| E\left[\|\vec{y}_1\|^2 - \|\vec{y}_2\|^2 \right] \right|$$

Where $\vec{y}_i = W^T \vec{x}_i$, $i = 1, 2$; U is the transformation matrix and $\vec{x}_i$ denotes n-dimensional random vector of class i.

$$\text{Let } D = \left| E\left[\|\vec{y}_1\|^2 - \|\vec{y}_2\|^2 \right] \right|$$

Substituting $\vec{y}_i = W^T \vec{x}_i$, $i = 1, 2$ in the above equation, we will get

$$D = \left| \sum_{i=1}^{k} \vec{w}_i^T M \vec{w}_i \right|.$$

Where $M = R_{x1} - R_{x2}$ and $R_{xi} = E\left[\vec{x}_i \vec{x}_i^T \right] \in \mathbb{R}^{n \times n}$, $i = 1, 2$.

Note: M is neither semi-positive nor semi-negative definite.

Basis of EST

Let $\{(\lambda_1^{(p)}, \vec{w}_1^{(p)}), (\lambda_2^{(p)}, \vec{w}_2^{(p)}), \ldots, (\lambda_{n_1}^{(p)}, \vec{w}_{n_1}^{(p)})\}$ and $\{(\lambda_1^{(n)}, \vec{w}_1^{(n)}),$ $(\lambda_2^{(n)}, \vec{w}_2^{(n)}), \ldots, (\lambda_{n_2}^{(n)}, \vec{w}_{n_2}^{(n)})\}$ be set of positive and negative eigenvalues, and corresponding eigenvectors respectively.

Basis of EST

Let $\{(\lambda_1^{(p)}, \vec{w}_1^{(p)}), (\lambda_2^{(p)}, \vec{w}_2^{(p)}), \ldots, (\lambda_{n_1}^{(p)}, \vec{w}_{n_1}^{(p)})\}$ and $\{(\lambda_1^{(n)}, \vec{w}_1^{(n)}), (\lambda_2^{(n)}, \vec{w}_2^{(n)}), \ldots, (\lambda_{n_2}^{(n)}, \vec{w}_{n_2}^{(n)})\}$ be set of positive and negative eigenvalues, and corresponding eigenvectors respectively. Then

1. if $\sum_{i=1}^{n_1} \lambda_i^{(p)} > \sum_{i=1}^{n_2} \lambda_i^{(n)}$ then basis of EST is $\{\vec{w}_1^{(p)}, \vec{w}_2^{(p)}, \ldots, \vec{w}_{n_1}^{(p)}\}$, *i.e.*

$$W = [\vec{w}_1^{(p)}, \vec{w}_2^{(p)}, \ldots, \vec{w}_{n_1}^{(p)}].$$

Basis of EST

Let $\{(\lambda_1^{(p)}, \vec{w}_1^{(p)}), (\lambda_2^{(p)}, \vec{w}_2^{(p)}), \ldots, (\lambda_{n_1}^{(p)}, \vec{w}_{n_1}^{(p)})\}$ and $\{(\lambda_1^{(n)}, \vec{w}_1^{(n)}),$ $(\lambda_2^{(n)}, \vec{w}_2^{(n)}), \ldots, (\lambda_{n_2}^{(n)}, \vec{w}_{n_2}^{(n)})\}$ be set of positive and negative eigenvalues, and corresponding eigenvectors respectively. Then

1. if $\sum_{i=1}^{n_1} \lambda_i^{(p)} > \sum_{i=1}^{n_2} \lambda_i^{(n)}$ then basis of EST is $\{\vec{w}_1^{(p)}, \vec{w}_2^{(p)}, \ldots, \vec{w}_{n_1}^{(p)}\}$, *i.e.*

$$W = [\vec{w}_1^{(p)}, \vec{w}_2^{(p)}, \ldots, \vec{w}_{n_1}^{(p)}].$$

2. if $\sum_{i=1}^{n_1} \lambda_i^{(p)} < \sum_{i=1}^{n_2} \lambda_i^{(n)}$ then basis of EST is $\{\vec{w}_1^{(n)}, \vec{w}_2^{(n)}, \ldots, \vec{w}_{n_2}^{(n)}\}$, *i.e.*

$$W = [\vec{w}_1^{(n)}, \vec{w}_2^{(n)}, \ldots, \vec{w}_{n_2}^{(n)}].$$

Basis of EST

Let $\{(\lambda_1^{(p)}, \vec{w}_1^{(p)}), (\lambda_2^{(p)}, \vec{w}_2^{(p)}), \ldots, (\lambda_{n_1}^{(p)}, \vec{w}_{n_1}^{(p)})\}$ and $\{(\lambda_1^{(n)}, \vec{w}_1^{(n)}), (\lambda_2^{(n)}, \vec{w}_2^{(n)}), \ldots, (\lambda_{n_2}^{(n)}, \vec{w}_{n_2}^{(n)})\}$ be set of positive and negative eigenvalues, and corresponding eigenvectors respectively. Then

1. if $\sum_{i=1}^{n_1} \lambda_i^{(p)} > \sum_{i=1}^{n_2} \lambda_i^{(n)}$ then basis of EST is $\{\vec{w}_1^{(p)}, \vec{w}_2^{(p)}, \ldots, \vec{w}_{n_1}^{(p)}\}$, *i.e.*

$$W = [\vec{w}_1^{(p)}, \vec{w}_2^{(p)}, \ldots, \vec{w}_{n_1}^{(p)}].$$

2. if $\sum_{i=1}^{n_1} \lambda_i^{(p)} < \sum_{i=1}^{n_2} \lambda_i^{(n)}$ then basis of EST is $\{\vec{w}_1^{(n)}, \vec{w}_2^{(n)}, \ldots, \vec{w}_{n_2}^{(n)}\}$, *i.e.*

$$W = [\vec{w}_1^{(n)}, \vec{w}_2^{(n)}, \ldots, \vec{w}_{n_2}^{(n)}].$$

3. if $\sum_{i=1}^{n_1} \lambda_i^{(p)} = \sum_{i=1}^{n_2} \lambda_i^{(n)}$ then basis of EST is $\{\vec{w}_1^{(p)}, \vec{w}_2^{(p)}, \ldots, \vec{w}_{n_1}^{(p)}\}$ or $\{\vec{w}_1^{n}, \vec{w}_2^{n}, \ldots, \vec{w}_{n_2}^{n}\}$, *i.e.*

$$W = [\vec{w}_1^{(p)}, \vec{w}_2^{(p)}, \ldots, \vec{w}_{n_1}^{(p)}] \text{ or } W = [\vec{w}_1^{(n)}, \vec{w}_2^{(n)}, \ldots, \vec{w}_{n_2}^{(n)}].$$

Disadvantages

- EST doesn't consider intra-class information of data patterns.

Disadvantages

- EST doesn't consider intra-class information of data patterns.

- The number of features to be generated is not user-specified. This may lead to poor generalization of a classifier.

Enhanced Eigenspace Separation Transform (EEST)

EEST consider intra-class distances along with the difference in the average lengths of the vectors in the two classes.

Intra-class distance in the feature space is defined as,

$$D_i = \frac{1}{N_i} \sum_{k=1}^{N_i} (\mathbf{y_i^{(k)}} - \bar{\mathbf{y}}_\mathbf{i})^\top (\mathbf{y_i^{(k)}} - \bar{\mathbf{y}}_\mathbf{i}), \; i = 1, 2.$$

Objective

$$\max_{W} \quad \left| E\left[||\mathbf{y_1}||^2 - ||\mathbf{y_2}||^2 \right] \right| - \sum_{i=1}^{2} D_i$$

Using,

$$I = I_1 + I_2,$$

$$I_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \left[(\mathbf{x_i^{(k)}} - \bar{\mathbf{x}}_\mathbf{i})(\mathbf{x_i^{(k)}} - \bar{\mathbf{x}}_\mathbf{i})^\top \right] \tag{1}$$

and

$$\vec{y}_i = W^T \vec{x}_i, \;\; i = 1, 2,$$

the objective can be reformulated as,

Objective

$$\max_W \left| \sum_{j=1}^{n} \mathbf{w_j}^\top M \mathbf{w_j} \right| - \sum_{j=1}^{n} \mathbf{w_j}^\top I \mathbf{w_j}$$

$$\textit{subject to } W^\top W = \mathbf{Id}$$

If we assume that $M \neq 0$, then there is at least one non-zero eigenvalue.

Let $\{(\lambda_1^{(P)},\ \mathbf{z_1^{(P)}}),\ (\lambda_2^{(P)},\ \mathbf{z_2^{(P)}}), \ldots,\ (\lambda_{n_1}^{(P)},\ \mathbf{z_{n_1}^{(P)}})\}$ be set of eigenvectors corresponding to positive set of eigenvalues of M.

Let $\{(\lambda_1^{(N)},\ \mathbf{z_1^{(N)}}),\ (\lambda_2^{(N)},\ \mathbf{z_2^{(N)}}), \ldots,\ (\lambda_{n_2}^{(N)},\ \mathbf{z_{n_2}^{(N)}})\}$ be set of eigenvectors corresponding to negative set of eigenvalues of M $(n_1 + n_2 \leq n)$. Let

$$M^{(P)} = \sum_{i=1}^{n_1} \lambda_i^{(P)} \mathbf{z_i^{(P)}} \left(\mathbf{z_i^{(P)}}\right)^\top, \tag{2}$$

$$M^{(N)} = \sum_{i=1}^{n_2} \left|\lambda_i^{(N)}\right| \mathbf{z_i^{(N)}} \left(\mathbf{z_i^{(N)}}\right)^\top, \tag{3}$$

Theorem

Let $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\} \subset \mathbb{R}^n$ *be an arbitrary set of vectors and* $U = (\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n})$. *Then the following statement hold good.*

$$f(U) \leq \max\left\{tr\left(U^\top (M^{(P)} - I)U\right), tr\left(U^\top (M^{(N)} - I)U\right)\right\},$$

where 'tr' denotes trace of a matrix.

Corollary

Let $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\} \subset \mathbb{R}^n$ *be an orthonormal solution basis for the criterion of EEST (i.e.* $\max_U f$*) and* $U = (\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n})$*, then the following statement hold good.*

$$f(U) \leq \max\left\{tr\left(U^\top (M^{(P)} - I)U\right), tr\left(U^\top (M^{(N)} - I)U\right)\right\}.$$

Theorem

The following cases hold good.

1. *Let* $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\} \subset \mathbb{R}^n$ *be an orthonormal set of eigenvectors of* $(M^{(P)} - I)$. *Let* $U = (\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n})$. *If* M *has no negative eigenvalues then*

$$\max_W f(W) = trace\left(U^\top \left(M^{(P)} - I\right) U\right)$$

   and $W = U$.

2. *Let* $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\} \subset \mathbb{R}^n$ *be an orthonormal set of eigenvectors of* $(M^{(N)} - I)$. *Let* $V = (\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n})$. *If* M *has no positive eigenvalues then*

$$\max_W f(W) = trace\left(V^\top \left(M^{(N)} - I\right) V\right)$$

   and $W = V$.

But M may not have either no negative or no positive eigenvalues always. So we have used the eigenspace of $(M^{(P)} - I)$ or $(M^{(N)} - I)$, thereby a major simplification of the decision surface is expected.

EEST Algorithm

1. Compute the $n \times n$ correlation difference matrix M

$$M = \frac{1}{N_1} \sum_{i=1}^{N_1} \mathbf{x}_1^{(\mathbf{i})} (\mathbf{x}_1^{(\mathbf{i})})^\top - \frac{1}{N_2} \sum_{j=1}^{N_2} \mathbf{x}_2^{(\mathbf{j})} (\mathbf{x}_2^{(\mathbf{j})})^\top .$$

EEST Algorithm

1. Compute the $n \times n$ correlation difference matrix M

$$M = \frac{1}{N_1}\sum_{i=1}^{N_1} \mathbf{x}_1^{(\mathbf{i})}(\mathbf{x}_1^{(\mathbf{i})})^\top - \frac{1}{N_2}\sum_{j=1}^{N_2} \mathbf{x}_2^{(\mathbf{j})}(\mathbf{x}_2^{(\mathbf{j})})^\top.$$

2. Compute the $n \times n$ intra-class distance matrix I using the equation (1).

EEST Algorithm

1. Compute the $n \times n$ correlation difference matrix M

$$M = \frac{1}{N_1}\sum_{i=1}^{N_1} \mathbf{x}_1^{(\mathbf{i})}(\mathbf{x}_1^{(\mathbf{i})})^\top - \frac{1}{N_2}\sum_{j=1}^{N_2} \mathbf{x}_2^{(\mathbf{j})}(\mathbf{x}_2^{(\mathbf{j})})^\top.$$

2. Compute the $n \times n$ intra-class distance matrix I using the equation (1).
3. Calculate the matrices $M^{(P)}$ and $M^{(N)}$ using the formulas in equations (2) and (3).

EEST Algorithm

1. Compute the $n \times n$ correlation difference matrix M

$$M = \frac{1}{N_1}\sum_{i=1}^{N_1} \mathbf{x_1^{(i)}}(\mathbf{x_1^{(i)}})^\top - \frac{1}{N_2}\sum_{j=1}^{N_2} \mathbf{x_2^{(j)}}(\mathbf{x_2^{(j)}})^\top.$$

2. Compute the $n \times n$ intra-class distance matrix I using the equation (1).
3. Calculate the matrices $M^{(P)}$ and $M^{(N)}$ using the formulas in equations (2) and (3).
4. Compute eigenvectors $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ of $M^{(P)} - I$ and $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ of $M^{(N)} - I$.

EEST Algorithm

1. Compute the $n \times n$ correlation difference matrix M

$$M = \frac{1}{N_1} \sum_{i=1}^{N_1} \mathbf{x_1^{(i)}} (\mathbf{x_1^{(i)}})^\top - \frac{1}{N_2} \sum_{j=1}^{N_2} \mathbf{x_2^{(j)}} (\mathbf{x_2^{(j)}})^\top .$$

2. Compute the $n \times n$ intra-class distance matrix I using the equation (1).
3. Calculate the matrices $M^{(P)}$ and $M^{(N)}$ using the formulas in equations (2) and (3).
4. Compute eigenvectors $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ of $M^{(P)} - I$ and $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ of $M^{(N)} - I$.
5. If $tr\left(U^\top (M^{(P)} - I) U\right) > tr\left(V^\top (M^{(N)} - I) V\right)$ then $W = U$ else $W = V$, where $U = (\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n})$, $V = (\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n})$

Experimental Results

Data Sets	Average Test Accuracy (%)			
	EEST	**EST**	**MMC**	**LDA**
Iris	95.289	93.225	89.848	**95.981**
N=150	**96.124**	93.884	89.467	95.733
D=4	**96.47**	95.046	89.864	95.456
C=2	**96.429**	95.095	89	96.286
Glass	**88.175**	86.322	59.544	57.181
N=214	**89.203**	87.374	62.729	59.421
D=10	**89.734**	88.313	63.906	58.969
C=7	**90.111**	89.984	66.191	62.191
Wine	95.973	88.565	96.323	**96.436**
N=178	96.457	89.146	**97.371**	97.730
D=13	96.629	90.05	**98.453**	98.189
C=3	97.098	90.078	**98.471**	98.235
Red Wine	**84.772**	84.118	53.096	53.955
N=1599	**86.106**	85.31	57.106	57.279
D=11	**87.317**	86.708	59.432	59.603
C=6	**88.201**	87.725	61.887	63.12

Continued...

Data Sets	Average Test Accuracy (%)			
	EEST	**EST**	**MMC**	**LDA**
White Wine	**86.608**	85.774	49.348	51.443
N=4898	**88.019**	87.148	54.387	55.965
D=11	**89.202**	88.262	58.746	60.161
C=7	**90.287**	89.29	62.323	62.81
Vehicle	**83.458**	80.344	71.449	73.51
N=946	**84.261**	80.838	72.109	74.676
D=18	**84.781**	82.413	71.668	75.439
C=4	**84.946**	83.054	70.9764	74.8814
Image Segmentation	**98.139**	97.671	94.104	-
N=2310	**98.611**	98.196	94.935	-
D=19	**98.839**	98.45	95.512	-
C=7	**99.004**	98.614	95.722	-
Multiple	**99.445**	94.635	97.176	-
N=2000	**99.541**	96.464	97.632	63.702
D=649	**99.601**	97.359	97.853	68.883
C=10	**99.64**	97.861	98.02	73.407

Continued...

Data Sets	Average Test Accuracy (%)			
	EEST	**EST**	**MMC**	**LDA**
Letter Recognition	**99.347**	98.059	82.494	86.629
N=20000	**99.511**	98.302	84.169	87.994
D=16	**99.56**	98.463	85.872	89.348
C=26	**99.654**	98.562	86.895	90.14
Yale Faces	**94.893**	86.928	83.13	52.696
N=165	**96.114**	89.691	88.293	49.512
D=2500	**96.313**	90.299	91.429	52.653
C=15	**96.083**	90.292	92.5	48.125
ORL Faces	**99.038**	89.507	90.536	68.929
N=400	**99.604**	94.568	95.95	69.9
D=2576	**99.796**	96.669	97.25	63.667
C=40	**99.84**	97.853	97.692	68.718

THANK YOU