Subspace Classifiers

Robert M. Haralick

Computer Science, Graduate Center
City University of New York
Euclidean Distance

Definition

The Squared Euclidean Distance Between $x$ and $y$ is defined by

$$||x - y||^2 = (x - y)'(x - y)$$

It is called the $L_2$ norm.
The squared $L_2$ norm of $x - y$ with respect to a symmetric positive definite matrix $A$ is given by

$$
\|x - y\|_A^2 = (x - y)'A(x - y)
$$
Definition

A square matrix $T$ is orthonormal if and only if

- Each column of $T$ has norm 1
- Every pair of different columns of $T$ is orthogonal
The Eigen Decomposition of a real square matrix $A$ is given by

$$A = T\Lambda T'$$

where $T$ is an orthonormal matrix and $\Lambda$ is a diagonal matrix.

- The columns of $T$ are the eigenvectors
- The diagonals of $\Lambda$ are the eigenvalues
- The $i^{th}$ column of $T$ and the $i^{th}$ diagonal element of $\Lambda$ constitute an eigenvector-eigenvalue pair

If $t_i$ is the $i^{th}$ column of $T$ and $\lambda_i$ is the $i^{th}$ eigenvalue of $\Lambda$, then

$$At_i = \lambda_i t_i$$
**Definition**

$t$ is an eigenvector of $A$ and $\lambda$ is the corresponding eigenvalue if and only if

$$At = \lambda t$$
Suppose that an $N \times N$ matrix $A = T \Lambda T$. Then the $i^{th}$ column of $T$, $t_i$, and the $i^{th}$ diagonal element, $\lambda_i$, of $\Lambda$ constitute an eigenvector eigenvalue pair.

**Proof**

Since $A = T \Lambda T'$, we can write

$$A t_i = T \Lambda T' t_i$$

$$= T \Lambda \begin{pmatrix} t'_1 \\ t'_2 \\ \vdots \\ t'_N \end{pmatrix} t_i$$
But every pair of different columns of $A$ are orthogonal

$$At_i = T\Lambda T' t_i$$

$$= T\Lambda(0, 0, \ldots, 1, \ldots 0)'$$

$$= T \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \lambda_i \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$At_i = \lambda_i t_i$$
Definition

A square matrix $A$ is called positive definite if and only if all its eigenvalues are positive.
The $L_2$ norm of $(x - y)$ with respect to positive definite square matrix $A$ is

$$(x - y)'A(x - y) = (x - y)' T \Lambda \frac{1}{2} T'(x - y)$$
$$= (x - y)' T \Lambda \frac{1}{2} \Lambda \frac{1}{2} T'(x - y)$$
$$= [(x - y)' T \Lambda \frac{1}{2}] [\Lambda \frac{1}{2} T'(x - y)]$$
$$= [\Lambda \frac{1}{2} T'(x - y)]' [\Lambda \frac{1}{2} T'(x - y)]$$
$$= \| [\Lambda \frac{1}{2} T'(x - y)] \|^2$$

This has a geometric meaning. An orthonormal matrix is either a rotation matrix or a rotation matrix with a reflection. So $(x - y)$ gets rotated by $T'$ and scaled by $\Lambda \frac{1}{2}$. After rotating and scaled, the norm is the standard $L_2$ norm (with respect to the identity matrix).
Mahalanobis Distance

Definition

The Mahalanobis distance between \( x \) and \( y \) with respect to covariance matrix \( \Sigma \) is defined by

\[
D(x - y) = \sqrt{(x - y)\Sigma^{-1}(x - y)}
\]

If \( \Sigma \) has the Eigen Decomposition \( \Sigma = T\Lambda T' \), then \( \Sigma^{-1} \) has the Eigen Decomposition \( \Sigma^{-1} = T\Lambda^{-1}T' \), where

\[
\Lambda^{-1} = \text{Diag} \left( \frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_N^2} \right) \quad \text{and} \quad \Lambda^{-\frac{1}{2}} = \text{Diag} \left( \frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_N} \right)
\]

Hence,

\[
D(x - y) = \sqrt{||\Lambda^{-\frac{1}{2}}T'(x - y)||^2}
\]

and means rotate \((x - y)\) by \( T' \) and then normalize by the standard deviations in the rotated space.
The ellipse in standard rotation is given by

\[
\frac{(x - x_0)^2}{a^2} \frac{(y - y_0)^2}{b^2} = 1
\]

- The center of the ellipse is \((x_0, y_0)\)
- The leftmost point of the ellipse is at \((x_0 - a, y_0)\)
- The rightmost point of the ellipse is at \((x_0 + a, y_0)\)
- The extent of the ellipse axis along the x-axis is 2a
- The bottommost point of the ellipse is at \((x_0, y_0 - b)\)
- The topmost point of the ellipse is at \((x_0, y_0 + b)\)
- The extent of the ellipse along the y-axis is 2b
Mahalanobis Distance

\[(x - y)'A(x - y) = \theta\]

Specifies an ellipse

\[(x - y)'A(x - y) \leq \theta\]

Specifies the insides of an ellipse

\[(x - \mu)'\Sigma^{-1}(x - \mu) = \theta\]
\[(x - \mu)'T\Lambda T'(x - \mu) = \theta\]
\[||\Lambda^{-\frac{1}{2}}T'(x - \mu)||^2 = \theta\]
The Hyperellipsoid

\[ \| \Lambda^{-\frac{1}{2}} T'(x - \mu) \|^2 = 1 \]

- Is the equation of an hyperellipsoid
- Whose center is \( \mu \)
- Which has been rotated by \( T' \)
- And scaled by \( \Lambda^{-\frac{1}{2}} \)
- The \( n^{th} \) column of \( T \) is \( t_n \)
- The \( n^{th} \) component of \( \mu \) is \( \mu_n \)
- The \( n^{th} \) diagonal entry of \( \Lambda^{-\frac{1}{2}} \) is \( \frac{1}{\sigma_n} \)
- The maximum point of the ellipse in the \( t_n \) direction is \( \mu_n + \sigma_n \)
- The minimum point of the ellipse in the \( t_n \) direction is \( \mu_n - \sigma_n \)
- The extent of the ellipse in the \( t_n \) direction is \( 2\sigma_n \)
The Rotated Ellipsoid
The Gaussian Classifier

\[ \mu_1 \text{ mean of class 1} \]
\[ \mu_2 \text{ mean of class 2} \]
\[ \Sigma_1 \text{ covariance matrix of class 1} \]
\[ \Sigma_2 \text{ covariance matrix of class 2} \]

Then \( \sqrt{(x - \mu_1)'\Sigma_1^{-1}(x - \mu_1)} \) is the distance between \( x \) and the distribution with mean \( \mu_1 \) and covariance \( \Sigma_1 \).

When \( |\Sigma_1| = |\Sigma_2| \) and \( P(c^1) = P(c^2) \), then assign vector \( x \) to class \( c^1 \) when

\[ (x - \mu_1)'\Sigma_1^{-1}(x - \mu_1) < (x - \mu_2)'\Sigma_2^{-1}(x - \mu_2) \]

Else assign to class \( c^2 \)
The Fisher Linear Discriminant

\[ v = \Sigma_W^{-1}(\mu_1 - \mu_2) \]

Assign \( x \) to class 1 if

\[ v'x \geq \theta \]
\[ (\Sigma_W^{-1}(\mu_1 - \mu_2))'x \geq \theta \]
\[ (\mu_1 - \mu_2)'\Sigma_W^{-1}x \geq \theta \]
\[ (\mu_2 - \mu_1)'\Sigma_W^{-1}x < \theta \]

When \( \Sigma_1 = \Sigma_2 \), the Gaussian classifier is a linear classifier and identical to the Fisher Linear Discriminant Classifier since \( \Sigma_W = \Sigma_1 = \Sigma_2 \)
High Dimensional Spaces

- When the set of features becomes large
- There are dependencies between features
- Dependencies cause covariance matrices to be singular
The Gaussian classifier is not stable
The Fisher Linear Discriminant Classifier is not stable
The support of the class conditional density function is in a translated subspace
Regularize the covariance, for $\alpha > 0$

$$\Sigma \leftarrow \Sigma + \alpha I$$
The subspace classifier was introduced by Satoshi Watanabe.
It assumes that the covariance matrices are near singular.
Works in the dense subspaces.
The entropy of a $K$-dimensional $N(\mu, \Sigma)$ density is

$$H = \frac{K}{2} (1 + \log(2\pi)) + \frac{K}{2} \sum_{k=1}^{K} \log \lambda_k$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K$ are the eigenvalues of $\Sigma$. 
CLAFIC

- \( M \) classes
- \( L_m \) feature vectors from class \( c_m \)
- \( D - \text{dimensional} \)
  - \( x_1^m, \ldots, x_{L_m}^m \)
- \( N = \sum_{m=1}^{M} L_m \) Total number of vectors
- \( \mu \) Global mean
- \( y_k^m \) Transformed feature vectors

\[
\mu = \frac{1}{N} \sum_{m=1}^{M} \sum_{k=1}^{L_m} x_k^m
\]

\[
y_k^m = x_k^m - \mu
\]

\[ S_m = \frac{1}{L_m} \sum_{k=1}^{L_m} y_k^m (y_k^m)' \]

Eigenvalues of \( S_m \)
\[ \lambda_1^m \geq \lambda_2^m \geq \ldots \geq \lambda_D^m \]

Corresponding Eigenvectors \( t_1^m, \ldots, t_D^m \)

Given \( \sigma, 0 < \sigma < 1 \),

The \( J_m \) most important directions for class \( m \) are
\[ t_1^m, \ldots, t_{J_m}^m \]

where
\[ \frac{\sum_{j=1}^{J_m-1} \lambda_j^m}{\sum_{j=1}^{D} \lambda_j^m} < \sigma \leq \frac{\sum_{j=1}^{J_m} \lambda_j^m}{\sum_{j=1}^{D} \lambda_j^m} \]
Assign \( x \) to class \( c_m \) where

\[
J_m \sum_{j=1}^{J_m} \left( (t_j^m)'x \right)^2 \geq \sum_{j=1}^{J_k} \left( (t_j^k)'x \right)^2 , \quad k = 1, \ldots, M
\]
Orthogonal Projection Operator

Proposition

Let $T_m$ be a matrix whose columns are orthonormal.

$$
T^m = \begin{pmatrix}
  \vdots & \vdots & \cdots & \vdots \\
  t_1^m & t_2^m & \cdots & t_{jm}^m \\
  \vdots & \vdots & \cdots & \vdots 
\end{pmatrix}
$$

$$
P_m = T^m(T^m)'
$$

Then $P_m$ is the orthogonal projection operator onto the subspace spanned by Col($T^m$)

Proof.

$$
P_m P_m = [T^m(T^m)'][T^m(T^m)'] = T^m[(T^m)'(T^m)](T^m)'
$$

$$
= T^m(T^m)' = P_m
$$

$$
P_m' = [T^m(T^m)']' = T^m(T^m)' = P_m
$$
Assign $x$ to class $c_m$ where

$$||P_m x||^2 \geq ||P_j x||^2, j = 1, \ldots, M$$

This is equivalent to
Assign $x$ to class $c_m$ where

$$x' P_m x \geq x' P_j x, j = 1, \ldots, M$$
Two Class Case

Use a threshold $\theta$

Assign $x$ to class $c_1$ if

$$\frac{x' P_1 x}{x' P_2 x} > \theta$$

Else assign $x$ to class $c_2$
Angle Between $x$ and a Subspace

Let $P$ be an orthogonal projection operator to a subspace $V$
Let $\theta$ be the angle between $x$ and $V$
Then
\[ \cos^2 \theta = \frac{x'Px}{x'x} \]
Assign $x$ to class $c_m$ when
\[ x'P_m x \geq x'P_j x, \ j = 1, \ldots, M \]
is the equivalent to Assign $x$ to class $c_m$ when
\[ \frac{x'P_m x}{x'x} \geq \frac{x'P_j x}{x'x}, \ j = 1, \ldots, M \]
\[ \cos^2 \theta_m \geq \cos^2 \theta_j, \ j = 1, \ldots, M \]
\[ \theta_m \leq \theta_j, \ j = 1, \ldots, M \]
Proposition

Let $P_1$ the orthogonal projection operator to subspace $S_1$, and $P_2$ the orthogonal projection operator to subspace $S_2$. If $S_1 \subseteq S_2$, then $P_1P_2 = P_2P_1 = P_1$

Proof.

Since $S_1 \subseteq S_2$, $S_2^\perp \subseteq S_1^\perp$. Let $x$ be an arbitrary vector. Then $x = u + v + w$ where $u \in S_1$, $v \in S_2 \cap S_1^\perp$, $w \in S_2^\perp$.

\[
P_1P_2x = P_1P_2(u+v+w) = P_1P_2u + P_1P_2v + P_1P_2w = P_1u + P_1v + 0 = u + 0 = u
\]

\[
P_2P_1x = P_2P_1(u+v+w) = P_2P_1u + P_2P_1v + P_2P_1w = P_2u + 0 + 0 = u
\]

\[
P_1x = P_1(u+v+w) = P_1u + P_1v + P_1w
\]

Therefore $P_1P_2 = P_2P_1 = P_1$
Let $U$ and $V$ be subspaces. Then the Direct Sum of $U$ and $V$ is denoted by $U \oplus V$ and is defined by

$$U \oplus V = \{ x \mid \text{for some } u \in U, v \in V, x = u + v \}$$
Orthogonal Projection Operators

**Proposition**

Let $P$ and $Q$ be orthogonal projection operators. If $PQ = QP$, then $PQ$ is an orthogonal projection operator onto $\text{Col}(P) \cap \text{Col}(Q)$

**Proof.**

Suppose $PQ = QP$. Then

\[
(PQ)(PQ) = P(QP)Q = P(PQ)Q = (PP)(QQ) = PQ
\]

Since $P$ and $Q$ are orthogonal projection operators, $P = P'$ and $Q = Q'$. Hence,

\[
(PQ)' = Q'P' = QP = PQ
\]
Proof.

Let $x \in \mathbb{R}^N$. Then $PQx \in \text{Col}(P)$ and $QPx \in \text{Col}(Q)$. Since $PQ = QP$, $PQx \in \text{Col}(P) \cap \text{Col}(Q)$.

Let $x \in \text{Col}(P) \cap \text{Col}(Q)$. Then $Px = x$ and $Qx = x$. This implies that $PQx = Px = x$. Hence every element of $\text{Col}(P) \cap \text{Col}(Q)$ is left invariant under the operator $PQ$. Let $y \in \text{Col}(P) \perp$. Then $QPy = Q0 = 0$. Likewise if $y \in \text{Col}(Q)$, $PQy = P0 = 0$.

Let $u \in \text{Col}(P) \perp$ and $v \in \text{Col}(Q) \perp$.

\[
PQ(u + v) = PQu + PQv = PQu
\]
\[
= QPu = 0
\]

So $y \in \text{Col}(P) \perp \oplus \text{Col}(Q) \perp$ implies $PQy = 0$. 

\[
\square
\]
Let \( P \) be an orthogonal projection operator to subspace \( V \) and \( Q \) be an orthogonal projection operator to subspace \( W \). If \( PQ = QP \), then \( P - PQ \) is the orthogonal projection operator to subspace \( V \cap W^\perp \)

**Proof.**

\[
(P - PQ)(P - PQ) = PP - PPQ - PQP + PQPPQ
\]
\[
= P - PQ - QPP + PPQQ
\]
\[
= P - PQ - QP + PQ
\]
\[
= P - PQ - PQ + PQ = P - PQ
\]

\[
(P - PQ)' = P' - (PQ)' = P' - Q'P'
\]
\[
= P - QP = P - PQ
\]
Proof. Let $x$ be arbitrary. Then $x = u + v + w$, where $u \in V \cap W^\perp$, $v \in V \cap W$, and $w \in V^\perp$.

\[
(P - PQ)x = (P - PQ)(u + v + w)
= u + v + 0 - PQu - PQv - PQw
= u + v - 0 - Pv - 0
= u + v - v = u
\]
Proposition

Let $P$ be an orthogonal projection operator to subspace $U$ and $Q$ be an orthogonal projection operator to subspace $V$. Then, $U \perp V$ if and only if $PQ = 0$

Proof.

Suppose $U \perp V$. Then $u \in U$ and $v \in V$ imply $u'v = 0$. Since $P$ projects to $U$, each column of $P$ is in $U$. Since $P$ is an orthogonal projection operator $P = P'$. Hence every row of $P$ is in $U$. Since $Q$ is a projection operator to $V$, every column of $Q$ is in $V$. Any entry of $PQ$ is of the form $p'q$ where $p$ is a column of $P$ and $q$ is a column of $Q$. But $p \in U$ and $q \in V$ and $U \perp V$ implies $p'q = 0$. 
Orthogonal Projection Operators

**Proposition**

If $P$ and $Q$ are orthogonal projection operators, then $PQ = 0$ if and only if $QP = 0$

**Proof.**

Suppose $PQ = 0$. Then

\[
PQ = P'Q' = (QP)'
\]

Hence, $PQ = 0$ implies $(QP)' = 0$. And, $(QP)' = 0$ implies $QP = 0$. 

Proposition

Let $P$ be an orthogonal projection operator to subspace $U$ and $Q$ be an orthogonal projection operator to subspace $V$. If $U \perp V$ then $P + Q$ is the orthogonal projection operator to $U \oplus V$.

Proof.

Since $U \perp V$, $PQ = QP = 0$. Then,

\[
(P + Q)(P + Q) = PP + PQ + QP + QQ
\]

\[
= P + 0 + 0 + Q = P + Q
\]

\[
(P + Q)' = P' + Q'
\]

\[
= P + Q
\]
Orthogonality

Proposition

If \( x \perp y \) then \( \{ z \mid \text{for some } \alpha, \beta, z = \alpha x = \beta y \} = \{0\} \)

Proof.

If \( x = 0 \) or \( y = 0 \) then it is clearly true. Without loss of generality, suppose \( y \neq 0 \).

\[
\begin{align*}
\alpha x &= \beta y \\
(\beta y)' \alpha x &= (\beta y)'(\beta y) \\
\alpha \beta y' x &= \beta^2 ||y||^2 \\
0 &= \beta^2 ||y||^2
\end{align*}
\]

Since \( y \neq 0 \), \( ||y||^2 > 0 \). Hence \( \beta = 0 \). And this implies that

\[
\{ z \mid \text{for some } \alpha, \beta, z = \alpha x = \beta y \} = \{0\}
\]
Corollary

Let $A$ and $B$ be symmetric $N \times N$ matrices. If $\text{Col}(A) \perp \text{Col}(B)$, then $AB = BA = 0$

Proof.

Since $\text{Col}(A) \perp \text{Col}(B)$, $A'B = 0$. But $A = A'$. Therefore $AB = 0$. Similarly, $BA = 0$. 
Proposition

Let $A$ and $B$ be subspaces of a vector space $V$. Then

$$\left( A \cap (A \cap B)^\perp \right) \perp \left( B \cap (A \cap B)^\perp \right)$$

Proof.

$$\left( A \cap (A \cap B)^\perp \right) \cap \left( B \cap (A \cap B)^\perp \right) = \left( (A \cap B) \cap (A \cap B)^\perp \right) = \{0\}$$
Orthogonal Projection Operators

Proposition

Let $A$ and $B$ be $N \times N$ symmetric matrices satisfying $AB = BA$. If $\text{Col}(A) \cap \text{Col}(B) = \{0\}$, then $AB = 0$

Proof.

Let $x \in \mathbb{R}^N$. Since $AB = BA$, $ABx \in \text{Col}(A)$ and $ABx \in \text{Col}(B)$. This implies that $ABx \in \text{Col}(A) \cap \text{Col}(B)$. But $\text{Col}(A) \cap \text{Col}(B) = \{0\}$. Hence, $ABx = 0$. Now if for any matrix $C$, $Cx$ for every $x$, $C = 0$. Therefore, $AB = 0$. If $AB = 0$, and $A = A'$ then $\text{Col}(A) \perp \text{Col}(B)$ and hence $\text{Col}(A) \cap \text{Col}(B) = \{0\}$

Corollary

Let $P$ and $Q$ be orthogonal projection operators with $\text{Col}(P) \cap \text{Col}(Q) = \{0\}$. If $PQ = QP$, then $PQ = 0$
Proposition

Let $P$ and $Q$ be orthogonal projection operators and $V = \text{Col}(P) \cap \text{Col}(Q)$. If $PQ = QP$, then $\text{Col}(P) \cap V^\perp \perp \text{Col}(Q) \cap V^\perp$.

Proof.

Let $S$ be the orthogonal projection operator onto $V$, $R$ be the orthogonal projection operator onto $\text{Col}(P) \cap S^\perp$, and $T$ be the orthogonal projection operator onto $\text{Col}(Q) \cap S^\perp$. Then $P = S + R$, $Q = S + T$. $RS = SR = 0$ since $\text{Col}(S) \perp \text{Col}(R)$, and $ST = TS = 0$ since $\text{Col}(S) \perp \text{Col}(T)$. By construction, $\text{Col}(R) = \text{Col}(P) \cap V^\perp$ and $\text{Col}(T) = \text{Col}(Q) \cap V^\perp$.

\[
PQ = QP \\
(S + R)(S + T) = (S + T)(S + R) \\
SS + ST + RS + RT = SS + SR + TS + TR \\
RT = TR
\]

By the previous corollary, $\text{Col}(R) \cap \text{Col}(T) = \{0\}$. By the previous proposition, $RT = TR$, $R = R'$ and $\text{Col}(R) \cap \text{Col}(T) = \{0\}$ imply $RT = 0$. Since $R = R'$, this implies that $\text{Col}(R) \perp \text{Col}(T)$ and hence $\text{Col}(P) \cap V^\perp \perp \text{Col}(Q) \cap V^\perp$. 

\[\square\]
Norm Inequalities

\[ \| \sum_{j=1}^{J} A_j x \|^2 \leq \left( \sum_{j=1}^{J} \| A_j x \| \right)^2 \]

\[ \left( \sum_{j=1}^{J} \| A_j x \| \right)^2 \leq J \sum_{j=1}^{J} \| A_j x \|^2 \]

\[ \| x - Px \|^2 = \| x \|^2 - \| Px \|^2, \text{ for orthogonal projection operator } P \]
Iterated Products of orthogonal projection operators converge to an orthogonal projection operator that projects onto the subspace that is common to all the projection operators. The earliest result for more than two projection operators in the iterations is by Nakano and Kakutani who published in Japanese in 1940. However Halperin is more commonly known whose book appeared in 1962.

Iterated Products of Projection Operators

**Proposition**

Let $P_1, ..., P_N$ be $N$ orthogonal Projection operators. Let $T = P_1 P_2 \ldots P_N$. Then for any $x$, $\lim_{k \to \infty} \| T^k x - T^{k+1} x \|^2 = 0$.

**Proof.**

Let $Q_0 = I$ and $Q_j = P_j Q_{j-1}$ so that $Q_N = T$. Then,

$$
\| T^k x - T^{k+1} x \|^2 = \| \sum_{n=1}^{N-1} (Q_n T^k x - Q_{n+1} T^k x) \|^2 
$$

$$
\leq \left( \sum_{n=1}^{N-1} \| Q_n T^k x - Q_{n+1} T^k x \| \right)^2 
$$

$$
\leq N \sum_{n=1}^{N-1} \| Q_n T^k x - Q_{n+1} T^k x \|^2 
$$

$$
\leq N \sum_{n=0}^{N-1} \| (Q_n T^k x) - P_n (Q_n T^k x) \|^2 
$$

$$
\leq N \sum_{n=1}^{N-1} \| Q_n T^k x \|^2 - \| P_n (Q_n T^k x) \|^2 
$$

Proof.

\[ \| T^k x - T^{k+1} x \|_2^2 \leq N \sum_{n=1}^{N-1} \| Q_n T^k x \|_2^2 - \| P_n (Q_n T^k x) \|_2^2 \]

\[ \leq N \sum_{n=1}^{N-1} \| Q_n T^k x \|_2^2 - \| Q_{n+1} T^k x \|_2^2 \]

\[ \leq N \left( \| Q_0 T^k x \|_2^2 - \| Q_N T^k x \|_2^2 \right) \]

\[ \leq N \left( \| T^k x \|_2^2 - \| T^{k+1} x \|_2^2 \right) \]

Since for any projection operator \( P \), \( \| Px \| \leq \| x \| \), the sequence
\(< \| T^0 x \|_2^2, \| T x \|_2^2, \| T^2 x \|_2^2, \ldots, \| T^K x \|_2^2, \ldots >\) is a decreasing sequence. Furthermore, it is bounded below by zero. Therefore, it converges. Hence,

\[ \lim_{k \to \infty} \| T^k x \|_2^2 - \| T^{k+1} x \|_2^2 = 0 \]

And this implies that

\[ \lim_{k \to \infty} \| T^k x - T^{k+1} x \|_2^2 = 0 \]
Theorem

Let $P_1, \ldots, P_N$ be orthogonal projection operators onto subspaces $M_1, \ldots, M_N$, of a vector space $V$, respectively. Let $P$ be the orthogonal projection operator onto $M = \cap_{n=1}^{N} M_n$. Let $T = P_1 P_2 \ldots P_N$. Then $\lim_{k \to \infty} T^k = P$. 

Iterated Products of Projection Operators
Theorem

Let \( P_k, k = 1, \ldots, K \) be orthogonal projection operators to subspaces \( S_1, \ldots, S_K \). Let \( S = \cap_{k=1}^{K} S_k \). Let \( \Gamma = \sum_{k=1}^{K} a_k P_k \) where \( 0 < a_k < 1 \) and \( \sum_{k=1}^{K} a_k = 1 \). Then the orthogonal projection operator \( P \) onto \( S \) is given by \( P = T T' \), where the columns of \( T \) are the eigenvectors of \( \Gamma \) having eigenvalue 1.

Proof.

Suppose \( v \in S \). Then \( v \in \text{Col}(P_k), k = 1, \ldots, K \). Then

\[
\Gamma v = \sum_{k=1}^{K} a_k P_k v = \sum_{k=1}^{K} a_k v = v
\]

Proof.

Let \( \psi \) be an eigenvector of \( \Gamma \) with eigenvalue 1. Then,

\[
\psi = \Gamma \psi = \sum_{k=1}^{K} a_k P_k \psi = \sum_{k=1}^{K} a_k \psi_k
\]

If there exists any \( k \) such that \( P_k \psi = \psi_k \) where \( ||\psi_k|| < ||\psi|| \)

\[
||\psi|| = \left| \left| \sum_{k=1}^{K} a_k \psi_k \right| \right|
\]

\[
\leq \sum_{k=1}^{K} a_k ||\psi_k|| < \sum_{k=1}^{K} a_k ||\psi|| = ||\psi||
\]

If the columns of \( T \) are the eigenvectors with eigenvalue 1, then \( TT' \) is the orthogonal projection operator onto \( S \).

\( \square \)
Direct Sum of Subspaces

Theorem

Let $P_k, k = 1, \ldots, K$ be orthogonal projection operators to subspaces $S_1, \ldots, S_K$ of an $N$-dimensional vector space. Let $S = \bigoplus_{k=1}^K S_k$. Let $\Gamma = \sum_{k=1}^K a_k P_k$ where $0 < a_k < 1$ and $\sum_{k=1}^K a_k = 1$. Then the orthogonal projection operator $P$ onto $S$ is given by $P = I - TT'$, where the columns of $T$ are the eigenvectors of $\Gamma$ having eigenvalue 0.

Proof.

Let $T$ be a matrix whose columns are the eigenvectors of $\Gamma$ associated with eigenvalue 0. Without loss of generality we take these eigenvectors to be those indexed by $M, \ldots, N$. Then

$$\text{Col}(T) = \bigcap_{n=M}^N S_n^\perp$$

and $TT'$ is the orthogonal projection operator onto $\bigcap_{n=M}^N S_n^\perp$. Since

$$\bigoplus_{n=M}^N S_n = \left(\bigcap_{n=M}^N S_n^\perp\right)^\perp$$

$I - TT'$ is the orthogonal projection operator onto $\bigoplus_{n=M}^N S_n$. 

\qed
Let $P_m$, $m = 1, \ldots, M$ be the $M$ orthogonal projection operators for classes $c_1, \ldots, c_M$

Let them project on subspaces $S_1, \ldots, S_M$, respectively

Let $T_m = S_m \cap (\oplus_{k \neq m} S_k)$

Let $R_m$ be the orthogonal projection operator onto $S_m \cap T_m^\perp$

Then $Q_m = P_m - R_m$

Assign $x$ to class $c_m$ if $x'Q_m x \geq x'Q_k x$, $k = 1, \ldots, M$.

Using Covariance Matrix

Let \( \mu_m, \Sigma_m, \ m = 1, \ldots, M \) be the estimated mean and covariance matrices from the training data for class \( c_m \).

Let \( \lambda_1^m \geq \lambda_2^m \geq \ldots \lambda_D^m \) be the eigenvalues of \( \Sigma_m \). Given \( \sigma \), \( 0 < \sigma < 1 \), determine the number \( J_m \) of directions for class \( c_m \) by,

\[
\frac{\sum_{j=1}^{J_m-1} \lambda_j^m}{\sum_{j=1}^D \lambda_j^m} < \sigma \leq \frac{\sum_{j=1}^{J_m} \lambda_j^m}{\sum_{j=1}^D \lambda_j^m}
\]

Let \( P_m \) be the orthogonal projection operator onto the space spanned by the first \( J_m \) eigenvectors of \( \Sigma_m \).

Assign \( x \) to class \( c_m \) where

\[
\|(I - P_m)(x - \mu_m)\| \leq \|(I - P_j)(x - \mu_j)\|, \ j = 1, \ldots, M
\]

Local Subspace Classifier

- For each class $c_m$, $m=1, \ldots, M$
- Find the closest $D_m + 1$ vectors, $D_m < D$, to $x$ of the training set for class $c_m$
- Denote them by $\mu_{0m}, \ldots, \mu_{D_m m}$
- Form the basis $B_m = \{\mu_{1m} - \mu_{0m}, \ldots, \mu_{D_m m} - \mu_{0m}\}$
- Calculate the orthogonal projection operator $P_m$ onto the space spanned by $B_m$
- The linear manifold for class $c_m$ is $L_m = \{x \mid x = B_m \alpha + \mu_{0m} \text{ for some } \alpha\}$
- Assign $x$ to class $c_m$ when the projection to the orthogonal complement space of $L_m$ is smallest

$$\|(I - P_m)(x - \mu_{0m})\| < \|(I - P_j)(x - \mu_{0j})\|, \ j = 1, \ldots, M$$