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

L_2 norm with a symmetric positive definite matrix

Definition

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

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

Orthonormal

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

Eigen Decomposition

Definition

The Eigen Decomposition of a real square matrix A is given by

$$A = T \Lambda T'$$

where T is an orthonormal matrix and Λ is a diagonal matrix.

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

If t_i is the i^{th} column of T and λ_i is the i^{th} eigenvalue of Λ, then

$$A t_i = \lambda_i t_i$$

Eigenvector Eigenvalue

Definition

t is an eigenvector of A and λ is the corresponding eigenvalue if and only if

$$At = \lambda t$$

Eigenvector Eigenvalue

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, λ_i, of Λ constitute an eigenvector eigenvalue pair

Proof

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

$$\begin{aligned} At_i &= T\Lambda T' t_i \\ &= T\Lambda \begin{pmatrix} t_1' \\ t_2' \\ \vdots \\ t_N' \end{pmatrix} t_i \end{aligned}$$

But every pair of different columns of A are orthogonal

$$
\begin{aligned}
At_i &= T\Lambda T' t_i \\
&= T\Lambda(0, 0, \ldots, 1, \ldots 0)' \\
&= T \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \lambda_i \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\
At_i &= \lambda_i t_i
\end{aligned}
$$

Positive Definite Matrix

Definition

A square matrix A is called positive definite if and only if all its eigenvalues of positive

L_2 norm with respect to square matrix A

The L_2 norm of $(x-y)$ with respect to positive definite square matrix A is

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

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 Σ is defined by

$$D(x-y) = \sqrt{(x-y)'\Sigma^{-1}(x-y)}$$

If Σ has the Eigen Decomposition $\Sigma = T\Lambda T'$, then Σ^{-1} has the Eigen Decomposition $\Sigma^{-1} = T\Lambda^{-1}T'$, where $\Lambda^{-1} = Diag\left(\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_N^2}\right)$ and $\Lambda^{-\frac{1}{2}} = Diag\left(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_N}\right)$ 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.

Ellipse

Definition

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 elipse

$$(x-y)'A(x-y)\leq\theta$$

Specifies the insides of an ellipse

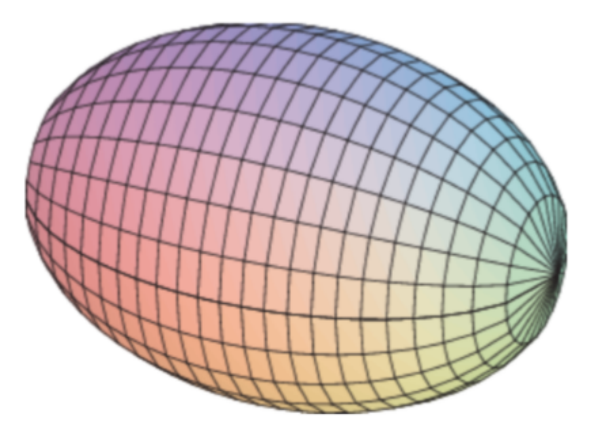
$$\begin{aligned}(x-\mu)'\Sigma^{-1}(x-\mu)&=\theta\\(x-\mu)'T\Lambda T'(x-\mu)&=\theta\\||\Lambda^{-\frac{1}{2}}T'(x-\mu)||^2&=\theta\end{aligned}$$

The Hyperellipsoid

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

- Is the equation of an hyperellipsoid
- Whose center is μ
- 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 μ is μ_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



μ_1	mean of class 1
μ_2	mean of class 2
Σ_1	covariance matrix of class 1
Σ_2	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 μ_1 and covariance Σ_1.

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

$$(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1) \quad < \quad (x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)$$

Else assign to class c^2

The Fisher Linear Discriminant

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

Assign x to class 1 if

$$\begin{aligned}
v'x &\geq \theta \\
\left(\Sigma_W^{-1}(\mu_1 - \mu_2)\right)'x &\geq \theta \\
(\mu_1 - \mu_2)'\Sigma_W^{-1\prime}x &\geq \theta \\
(\mu_1 - \mu_2)'\Sigma_W^{-1}x &\geq \theta \\
(\mu_2 - \mu_1)'\Sigma_W^{-1}x &< \theta
\end{aligned}$$

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

Singular Covariance Matrices

- 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 $\alpha > 0$

$$\Sigma \leftarrow \Sigma + \alpha I$$

Subspace Classifier

- The subspace classifier was introduced by Satosi Watanabe
- It assumes that the covariance matrices are near singular
- Works in the dense subspaces

Entropy Multivariate Gaussian Distribution

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 Σ

CLAFIC

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

$$
\begin{aligned}
\mu &= \frac{1}{N} \sum_{m=1}^{M} \sum_{k=1}^{L_m} x_k^m \\
y_k^m &= x_k^m - \mu
\end{aligned}
$$

Satosi Watanabe, *Feature Compression*, **Advances in Information Sciences**, Vol 3, Julius Tou (ed.), Plenun Press, New York, 1970, pp.63-111.

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

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 σ, $0 < \sigma < 1$,

The J_m most important directions for class m are

$$t_1^m, \ldots, t_{J_m}^m$$

where

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

Class Featuring Information Compression

Assign x to class c_m where

$$\sum_{j=1}^{J_m} \left((t_j^m)'x\right)^2 \geq \sum_{j=1}^{J_k} \left((t_j^k)'x\right)^2, k = 1, \ldots, M$$

Orthogonal Projection Operator

Proposition

Let T_m be a matrix whose columns are orthonormal.

$$\begin{aligned} T^m &= \begin{pmatrix} \vdots & \vdots & \dots & \vdots \\ t_1^m & t_2^m & \dots & t_{J_m}^m \\ \vdots & \vdots & \dots & \vdots \end{pmatrix} \\ P_m &= T^m(T^m)' \end{aligned}$$

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

Proof.

$$\begin{aligned} P_m P_m &= [T^m(T^m)'][T^m(T^m)'] = T^m[(T^m)'T^m](T^m)' \\ &= T^m(T^m)' = P_m \\ P_m' &= [T^m(T^m)']' = T^m(T^m)' = P_m \end{aligned}$$

Orthogonal Projection Operator

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 θ

Assign x to class c_1 if

$$\frac{x'P_1x}{x'P_2x} > \theta$$

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 θ be the angle between x and V
Then

$$cos^2\theta = \frac{x'Px}{x'x}$$

Assign x to class c_m when

$$x'P_mx \geq x'P_jx, \; j = 1, \ldots, M$$

is the equivalent to Assign x to class c_m when

$$\begin{aligned} \frac{x'P_mx}{x'x} &\geq \frac{x'P_jx}{x'x}, \; j = 1, \ldots, M \\ cos^2\theta_m &\geq cos^2\theta_j, \; j = 1, \ldots, M \\ \theta_m &\leq \theta_j, \; j = 1, \ldots, M \end{aligned}$$

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 $P_1 P_2 = P_2 P_1 = P_1$

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 $u \in S_1$, $v \in S_2 \cap S_1^{\perp}$, $w \in S_2^{\perp}$.

$$
\begin{aligned}
P_1 P_2 x &= P_1 P_2 (u + v + w) = P_1 P_2 u + P_1 P_2 v + P_1 P_2 w \\
&= P_1 u + P_1 v + 0 = u + 0 = u \\
P_2 P_1 x &= P_2 P_1 (u + v + w) = P_2 P_1 u + P_2 P_1 v + P_2 P_1 w \\
&= P_2 u + 0 + 0 = u \\
P_1 x &= P_1 (u + v + w) = P_1 u + P_1 v + P_1 w \\
&= u = 0 + 0 = u \\
\textit{Therefore} \quad & P_1 P_2 = P_2 P_1 = P_1
\end{aligned}
$$

□

Direct Sum

Definition

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 $Col(P) \cap Col(Q)$

Proof.

Suppose $PQ = QP$. Then

$$\begin{aligned}(PQ)(PQ) &= P(QP)Q = P(PQ)Q \\ &= (PP)(QQ) = PQ\end{aligned}$$

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

$$\begin{aligned}(PQ)' &= Q'P' = QP \\ &= PQ\end{aligned}$$

□

Proof Continued

Proof.

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

$$\begin{aligned} PQ(u+v) &= PQu + PQv = PQu \\ &= QPu = 0 \end{aligned}$$

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

□

Orthogonal Projection Operators

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.

$$
\begin{aligned}
(P - PQ)(P - PQ) &= PP - PPQ - PQP + PQPQ \\
&= P - PQ - QPP + PPQQ \\
&= P - PQ - QP + PQ \\
&= P - PQ - PQ + PQ = P - PQ \\
(P - PQ)' &= P' - (PQ)' = P' - Q'P' \\
&= P - QP = P - PQ
\end{aligned}
$$

□

Proof Continued

Proof.

Let x be arbitrary. Then $x = u + v + w$, where $u \in V \cap W^\perp$, $v \in V \cap W$, and $w \in V^\perp$.

$$
\begin{aligned}
(P - PQ)x &= (P - PQ)(u + v + w) \\
&= u + v + 0 - PQu - PQv - PQw \\
&= u + v - 0 - Pv - 0 \\
&= u + v - v = u
\end{aligned}
$$

□

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 $p'q = 0$. □

Proposition

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

Proof.

Suppose $PQ = 0$. Then

$$
\begin{aligned}
PQ &= P'Q' \\
&= (QP)'
\end{aligned}
$$

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

Orthogonal Projection Operators

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,

$$\begin{aligned} (P+Q)(P+Q) &= PP + PQ + QP + QQ \\ &= P + 0 + 0 + Q = P + Q \\ (P+Q)' &= P' + Q' \\ &= P + Q \end{aligned}$$

$\square$

Orthogonality

Proposition

If $x \perp y$ *then* $\{z \mid \textit{for some } \alpha, \beta, z = \alpha x = \beta y\} = \{0\}$

Proof.

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

$$\begin{aligned} \alpha x &= \beta y \\ (\beta y)'\alpha x &= (\beta y)'(\beta y) \\ \alpha\beta y' x &= \beta^2 ||y||^2 \\ 0 &= \beta^2 ||y||^2 \end{aligned}$$

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

$$\{z \mid \textit{for some } \alpha, \beta, z = \alpha x = \beta y\} = \{0\}$$

□

Corollary

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

Proof.

Since $Col(A) \perp Col(B)$, $A'B = 0$. 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.

$$\begin{aligned}\left(A \cap (A \cap B)^{\perp}\right) \cap \left(B \cap (A \cap B)^{\perp}\right) &= \left((A \cap B) \cap (A \cap B)^{\perp}\right) \\ &= \{0\}\end{aligned}$$

□

Orthogonal Projection Operators

Proposition

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

Proof.

Let $x \in \mathbb{R}^N$. Since $AB = BA$, $ABx \in Col(A)$ and $ABx \in Col(B)$. This implies that $ABx \in Col(A) \cap Col(B)$. But $Col(A) \cap Col(B) = \{0\}$. 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 $Col(A) \perp Col(B)$ and hence $Col(A) \cap Col(B) = \{0\}$ □

Corollary

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

Orthogonal Projection Operators

Proposition

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

Proof.

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

$$\begin{aligned} PQ &= QP \\ (S+R)(S+T) &= (S+T)(S+R) \\ SS + ST + RS + RT &= SS + SR + TS + TR \\ RT &= TR \end{aligned}$$

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

Norm Inequalities

$$
\begin{aligned}
||\sum_{j=1}^{J} A_j x||^2 &\leq \left(\sum_{j=1}^{J} ||A_j x|| \right)^2 \\
\left(\sum_{j=1}^{J} ||A_j x|| \right)^2 &\leq J \sum_{j=1}^{J} ||A_j x||^2 \\
||x - Px||^2 &= ||x||^2 - ||Px||^2, \text{for orthogonal projection operator } P
\end{aligned}
$$

Iterated Products of Projection Operators

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.

Anupan Netyanun and Donald Solmon, *Products of Projections in Hilbert Space*, **The American Mathematical Monthly**, Vol. 113, No. 7, 2006, 644-648.

Iterated Products of Projection Operators

Proposition

Let $P_1, ..., P_N$ be N orthogonal Projection operators. Let $T = P_1 P_2 \dots P_N$. 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,

$$
\begin{aligned}
||T^k x - T^{k+1}x||^2 &= ||\sum_{n=1}^{N-1}(Q_n T^k x - Q_{n+1} T^k x)||^2 \\
&\le \left(\sum_{n=1}^{N-1} ||Q_n T^k x - Q_{n+1} T^k x||\right)^2 \\
&\le N \sum_{n=0}^{N-1} ||Q_n T^k x - Q_{n+1} T^k x||^2 \\
&\le N \sum_{n=0}^{N-1} ||(Q_n T^k x) - P_n(Q_n T^k x)||^2 \\
&\le N \sum_{n=1}^{N-1} ||Q_n T^k x||^2 - ||P_n(Q_n T^k x)||^2
\end{aligned}
$$

Proof.

$$
\begin{aligned}
||T^k x - T^{k+1}x||^2 &\leq N\sum_{n=1}^{N-1} ||Q_n T^k x||^2 - ||P_n(Q_n T^k x)||^2 \\
&\leq N\sum_{n=1}^{N-1} ||Q_n T^k x||^2 - ||Q_{n+1} T^k x||^2 \\
&\leq N\left(||Q_0 T^k x||^2 - ||Q_N T^k x||^2\right) \\
&\leq N\left(||T^k x||^2 - ||T^{k+1}x||^2\right)
\end{aligned}
$$

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

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

And this implies that

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

□

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 $\lim_{k\to\infty} T^k = P$.

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 $P = TT'$, where the columns of T are the eigenvectors of Γ having eigenvalue 1.

Proof.

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

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

□

C.W. Therrien, *Eigenvalue Properties of Projection Operators and Their Application to the Subspace Method of Feature Extraction*, **IEEE Transactions on Computers**, September 1975, pp. 944-948.

Proof Continued

Proof.

Let ψ be an eigenvector of Γ 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 $||\psi_k|| < ||\psi||$

$$\begin{aligned} ||\psi|| &= ||\sum_{k=1}^{K} a_k \psi_k|| \\ &\leq \sum_{k=1}^{K} a_k ||\psi_k|| < \sum_{k=1}^{K} a_k ||\psi|| = ||\psi|| \# \end{aligned}$$

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

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 $S = \oplus_{k=1}^{K} S_k$. 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 Γ having eigenvalue 0.

Proof.

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

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

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

$$\oplus_{n=M}^{N} S_n = \left(\cap_{n=M}^{N} S_n^{\perp} \right)^{\perp}$$

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

Subtracting Overlapped Subspaces

- 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 $x' Q_m x \geq x' Q_k x$, $k = 1, \ldots, M$.

Satosi Watanabe, *Subspace Method in Pattern Recognition*, **Proceedings First International Joint Conference Pattern Recognition**, Washington D.C. 1973, pp. 25-32.

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 Σ_m. Given σ, $0 < \sigma < 1$, 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 Σ_m.

Assign x to class c_m where

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

Jorma Laaksonen and Erkki Oja, *Density Function Interpretation of Subspace Classification Methods*, **Proceedings of SCIA '97**, Lappenranta, Finland, June 1997, pp. 487-492.

Local Subspace Classifier

- For each class c_m, m=1,...,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 $L_m = \{x \mid x = B_m \alpha + \mu_{0m}$ for some $\alpha\}$
- Assign x to class c_m when the projection to the orthogonal complement space of L_m is smallest

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

Jorma Laaksonen, *Local Subspace Classifier*, **Proceedings of WSOM'97**, Espoo, Finland, June 1997, pp. 32-37.