coefficients of expansion or projection $x_i = \frac{v.\, v_i}{v_i.\, v_i}$

Determinants

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

$$det(AB) = det(A)det(B)$$

$$det(A+B) \neq det(A) + det(B)$$

$$det(A^{-1}) = \frac{1}{det(A)}$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

$$(A^T)^{-1} = (A^{-1})^T$$

Decomposition of Σ

$$\Sigma = \Phi\Lambda\Phi^{-1}$$

Eigenvalues and Eigenvectors

$$det(A - \lambda I) = 0$$

$$Av = \lambda v$$

$$\prod_i \lambda_i = det(A)$$

Variance

$$\sigma_X^2 = Var(X) = E((X-\mu_X)^2)$$

$$\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2$$

Covariance

$$\sigma_{XY} = Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y))$$

$$\hat{\sigma}_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y)$$

Gaussian Distribution

$$p(x) = N(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\exp^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$p(\mathbf{x}) = N(\mu, \Sigma) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}}\exp[-\frac{1}{2}(\mathbf{x}-\mu)^t\Sigma^{-1}(\mathbf{x}-\mu)]$$

Linear Transformations

$$\mathbf{Y} = A^t\mathbf{X}$$

$$p(\mathbf{y}) \sim N(A^t\mu_{\mathbf{X}}, A^t\Sigma_{\mathbf{X}}A)$$

Whitening Transformation

$$A_w = \Phi\Lambda^{-1/2} \qquad \mathbf{Y} = A_w^t\mathbf{X}$$

$$\Sigma_{\mathbf{X}} = \Phi\Lambda\Phi^{-1} \qquad p(\mathbf{y}) \sim N(A_w^t\mu_{\mathbf{X}}, I)$$

The overall risk

$$R = \int R(a(\mathbf{x})/\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

Discriminants

$$g_i(\mathbf{x}) = \ln p(\mathbf{x}/\omega_i) + \ln P(\omega_i)$$

average probability error

$$P(error) = \int_{-\infty}^{\infty} P(error, x)dx = \int_{-\infty}^{\infty} P(error/x)p(x)dx$$

Conditional Risk

$$R(a_i/\mathbf{x}) = \sum_{j=1}^{c}\lambda(a_i/\omega_j)P(\omega_j/\mathbf{x})$$

Discriminant (Multivariate Gaussian)

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\mu_i)^t\Sigma_i^{-1}(\mathbf{x}-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

Case 1 Discriminants

$$g_i(\mathbf{x}) = \mathbf{w}_i^t\mathbf{x} + w_{i0}$$

where $\mathbf{w}_i = \frac{1}{\sigma^2}\mu_i$, and $w_{i0} = -\frac{1}{2\sigma^2}\mu_i^t\mu_i + \ln P(\omega_i)$

$$\mathbf{w}^t(\mathbf{x} - \mathbf{x_0}) = 0$$

where $\mathbf{w} = \mu_i - \mu_j$, and $\mathbf{x}_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{\|\mu_i - \mu_j\|^2}\ln\frac{P(\omega_i)}{P(\omega_j)}(\mu_i - \mu_j)$

$$g_i(\mathbf{x}) = -\|\mathbf{x} - \mu_i\|^2$$

where $\|\mathbf{x} - \mu_i\|^2 = (\mathbf{x} - \mu_i)^t(\mathbf{x} - \mu_i)$

Case 2 Discriminants

$$g_i(\mathbf{x}) = \mathbf{w}_i^t\mathbf{x} + w_{i0}$$

(*linear discriminant*)

where $\mathbf{w}_i = \Sigma^{-1}\mu_i$, and $w_{i0} = -\frac{1}{2}\mu_i^t\Sigma^{-1}\mu_i + \ln P(\omega_i)$

$$\mathbf{w}^t(\mathbf{x} - \mathbf{x_0}) = 0$$

where $\mathbf{w} = \Sigma^{-1}(\mu_i - \mu_j)$ and $\mathbf{x}_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\ln[P(\omega_i)/P(\omega_j)]}{(\mu_i - \mu_j)^t\Sigma^{-1}(\mu_i - \mu_j)}(\mu_i - \mu_j)$

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)^t\Sigma^{-1}(\mathbf{x} - \mu_i)$$

Case 3 Discriminants

$$g_i(x) = x^tW_ix + w_i^tx + w_{i0}$$

(*quadratic discriminant*)

where $\mathbf{W}_i = -\frac{1}{2}\Sigma_i^{-1}$, $\mathbf{w}_i = \Sigma_i^{-1}\mu_i$, and $w_{i0} = -\frac{1}{2}\mu_i^t\Sigma_i^{-1}\mu_i - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$

$$\mathbf{x}^t(\mathbf{W}_1 - \mathbf{W}_2)\mathbf{x} + (\mathbf{w}_1^t - \mathbf{w}_2^t)\mathbf{x} + (w_{1,0} - w_{2,0}) = 0$$

ML

$$\sum_{k=1}^{n}\nabla_{\boldsymbol{\theta}}\ln p(\mathbf{x}_k/\boldsymbol{\theta}) = 0$$

$$\ln p(\mathbf{x}/\boldsymbol{\mu}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma|$$

$$\ln p(x_k/\boldsymbol{\theta}) = -\frac{1}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}(x_k - \mu)^2$$

Error Bounds

$$P(error) \le P^{\beta}(\omega_1)P^{1-\beta}(\omega_2)e^{-k(\beta)}$$

$$k(\beta) = \frac{\beta(1-\beta)}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t[(1-\beta)\boldsymbol{\Sigma}_1 + \beta\boldsymbol{\Sigma}_2]^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \frac{1}{2}\ln\frac{|(1-\beta)\boldsymbol{\Sigma}_1 + \beta\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|^{1-\beta}|\boldsymbol{\Sigma}_2|^{\beta}}.$$

Recursive Bayes Learning

$$p(\boldsymbol{\theta}/D^n) = \frac{p(\mathbf{x}_n/\boldsymbol{\theta})p(\boldsymbol{\theta}/D^{n-1})}{\int p(\mathbf{x}_n/\boldsymbol{\theta})p(\boldsymbol{\theta}/D^{n-1})d\boldsymbol{\theta}}$$

where $p(\boldsymbol{\theta}/D^0) = p(\boldsymbol{\theta})$

BE

$$p(\mathbf{x}/D) = \int p(\mathbf{x}/\boldsymbol{\theta})p(\boldsymbol{\theta}/D)d\boldsymbol{\theta} \qquad p(\boldsymbol{\theta}/D) = \frac{p(D/\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(D)} = a\prod_{k=1}^{n}p(\mathbf{x}_k/\boldsymbol{\theta})p(\boldsymbol{\theta})$$

MAP for MultV Gaussian Unknown θ=μ

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\mu}) \sim N(\boldsymbol{\mu_0}, \Sigma_{\mu} = \mathbf{I}\sigma_{\boldsymbol{\mu}_0})$$

$$p(\mathbf{x}/\boldsymbol{\theta}) = p(\mathbf{x}/\boldsymbol{\mu}) \sim N(\boldsymbol{\mu}, \Sigma = \mathbf{I}\sigma_{\boldsymbol{\mu}})$$

$$\hat{\boldsymbol{\mu}} = \frac{\boldsymbol{\mu_0} + \frac{\sigma_{\boldsymbol{\mu}_0}^2}{\sigma_{\boldsymbol{\mu}}^2}\sum_{k=1}^{n}\mathbf{x}_k}{1 + \frac{\sigma_{\boldsymbol{\mu}_0}^2}{\sigma_{\boldsymbol{\mu}}^2}n}$$

BE: UniV Gaussian, Unknown θ=μ, Known σ and p(μ)

$$p(x/D) \sim N(\mu_n, \sigma^2 + \sigma_n^2)$$

$$p(\mu/D) = c \times \frac{1}{\sqrt{2\pi}\sigma_n}\exp[-\frac{1}{2}(\frac{\mu - \mu_n}{\sigma_n})^2] \qquad \mu_n = (\frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2})\bar{x}_n + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}\mu_0 \qquad \sigma_n^2 = \frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}$$

BE: MultV Gaussian, Unknown θ=μ, Known Σ and p(μ)

$$\boldsymbol{\mu}_n = \Sigma_0(\Sigma_0 + \frac{1}{n}\Sigma)^{-1}\bar{\mathbf{x}}_n + \frac{1}{n}\Sigma(\Sigma_0 + \frac{1}{n}\Sigma)^{-1}\boldsymbol{\mu}_0$$

$$p(\mathbf{x}/D) = \int p(\mathbf{x}/\boldsymbol{\mu})p(\boldsymbol{\mu}/D)d\boldsymbol{\mu} \sim N(\boldsymbol{\mu}_n, \Sigma + \Sigma_n)$$

$$\Sigma_n = \Sigma_0(\Sigma_0 + \frac{1}{n}\Sigma)^{-1}\frac{1}{n}\Sigma$$

$$p(\boldsymbol{\mu}/D) = c \times \exp[-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^t\Sigma_n^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_n)]$$

PCA Formula

$$\overline{\mathbf{x}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_i \qquad \Phi_i = \mathbf{x}_i - \overline{\mathbf{x}} \qquad \mathbf{A=[\Phi_1\ \Phi_2\ \dots\ \Phi_M]}$$

$$\Sigma_{\mathbf{x}} = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T = \frac{1}{M}\sum_{i=1}^{M}\Phi_i\Phi_i^T = \frac{1}{M}AA^T$$

$$\mathbf{x} - \overline{\mathbf{x}} = \sum_{i=1}^{D} y_i u_i = y_1u_1 + y_2u_2 + \dots + y_Du_D$$

$$y_i = \frac{(\mathbf{x} - \overline{\mathbf{x}})^T u_i}{u_i^T u_i} = (\mathbf{x} - \overline{\mathbf{x}})^T u_i \qquad if\ \ \|u_i\| = 1$$

$$\frac{\sum_{i=1}^{K}\lambda_i}{\sum_{i=1}^{D}\lambda_i} > T \quad where\ T\ is\ a\ threshold\ (e.g., 0.9)$$

$$\hat{\mathbf{x}} = \sum_{i=1}^{K} y_i u_i + \overline{\mathbf{x}} = y_1u_1 + y_2u_2 + \dots + y_Ku_K + \overline{\mathbf{x}}$$

$$\|\mathbf{x} - \hat{\mathbf{x}}\| = \frac{1}{2}\sum_{i=K+1}^{D}\lambda_i$$

Linear Discriminants

$$g(\mathbf{x}) = \mathbf{w}^t\mathbf{x} + w_0 = \sum_{i=1}^{d} w_i x_i + w_0$$

$$J(\mathbf{w}, w_0) = \frac{1}{n}\sum_{k=1}^{n}[z_k - \hat{z}_k]^2$$

$$z_k = \begin{cases} +1\ if\ \ \mathbf{x}_k \in \omega_1 \\ -1\ if\ \ \mathbf{x}_k \in \omega_2 \end{cases}$$

$$\hat{z}_k = \begin{cases} +1\ if\ \ g(\mathbf{x}_k) > 0 \\ -1\ if\ \ g(\mathbf{x}_k) < 0 \end{cases}$$

$$r = \frac{g_i(\mathbf{x}) - g_j(\mathbf{x})}{\|\mathbf{w}_i - \mathbf{w}_j\|}$$

$$g(\mathbf{x}) = r\,\|\mathbf{w}\|$$

sample covariance matrix

$$\hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k - \hat{\boldsymbol{\mu}})(\mathbf{x}_k - \hat{\boldsymbol{\mu}})^t$$

$$H = \begin{bmatrix} \frac{\partial^2 J(\boldsymbol{\alpha})}{\partial\alpha_1^2} & \frac{\partial^2 J(\boldsymbol{\alpha})}{\partial\alpha_1\alpha_2} & \dots & \frac{\partial^2 J(\boldsymbol{\alpha})}{\partial\alpha_1\alpha_d} \\ \frac{\partial^2 J(\boldsymbol{\alpha})}{\partial\alpha_2\alpha_1} & \frac{\partial^2 J(\boldsymbol{\alpha})}{\partial\alpha_2^2} & \dots & \frac{\partial^2 J(\boldsymbol{\alpha})}{\partial\alpha_2\alpha_d} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 J(\boldsymbol{\alpha})}{\partial\alpha_d\alpha_1} & \frac{\partial^2 J(\boldsymbol{\alpha})}{\partial\alpha_d\alpha_2} & & \frac{\partial^2 J(\boldsymbol{\alpha})}{\partial\alpha_d^2} \end{bmatrix}$$

LDA Formula

$$M = \sum_{i=1}^{C} M_i \qquad \boldsymbol{\mu} = \frac{1}{C}\sum_{i=1}^{C}\boldsymbol{\mu}_i$$

$$S_w = \sum_{i=1}^{C}\sum_{j=1}^{M_i}(\mathbf{x}_{ij} - \boldsymbol{\mu}_i)(\mathbf{x}_{ij} - \boldsymbol{\mu}_i)^T$$

$$S_b = \sum_{i=1}^{C}(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$$

$$S_w^{-1}S_b u_k = \lambda_k u_k$$

Problem 3: Minimize $\frac{1}{2}\|w\|^2 + c\sum_{k=1}^{n}\psi_k$

subject to $z_k(w^t x_k + w_0) \ge 1 - \psi_k, \quad k = 1, 2, \dots, n$

Problem 4: Maximize $\sum_{k=1}^{n}\lambda_k - \frac{1}{2}\sum_{k,j}^{n}\lambda_k\lambda_j z_k z_j x_j^t x_k$

subject to $\sum_{k=1}^{n} z_k\lambda_k = 0$ and $0 \le \lambda_k \le c,\ k = 1, 2, \dots, n$

VC dimension → probabilistic upper bound on the generalization error of a classifier

$$error_{true} \le error_{train} + \sqrt{\frac{h(\log(2n/h) + 1) - \log(\delta/4)}{n}}$$

Kernels:

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x.y})^p \quad \text{or} \quad (\gamma(\mathbf{x.y}) + r)^p$$

$$K(\mathbf{x}, \mathbf{y}) = \tanh(\gamma(\mathbf{x.y}) + r)$$

$$K(\mathbf{x}, \mathbf{y}) = \exp(-\gamma\|\mathbf{x} - \mathbf{y}\|^2)$$

The SVM discriminant

$$g(\mathbf{x}) = \sum_{k=1}^{n} z_k\lambda_k(\mathbf{x}_k^t\mathbf{x}) + w_0 = \sum_{k=1}^{n} z_k\lambda_k(\mathbf{x}\,.\,\mathbf{x}_k) + w_0$$

$$g(\mathbf{x}) = \sum_{k=1}^{n} z_k\lambda_k(\boldsymbol{\Phi}(\mathbf{x})\,.\,\boldsymbol{\Phi}(\mathbf{x}_k)) + w_0$$

Naïve Bayesian Net

$$P(\omega_i / x_1, x_2, \dots, x_n) = \frac{p(x_1, x_2, \dots, x_n / \omega_i)P(\omega_i)}{p(x_1, x_2, \dots, x_n)}$$

$$where \quad p(x_1, x_2, \dots, x_n / \omega_i) = \prod_{j=1}^{n} p(x_j / \omega_i)$$

Markov property

$$p(x_1, x_2, \dots, x_n) = p(x_1/\pi_1)\,p(x_2/\pi_2)\dots p(x_n/\pi_n) = \prod_{i=1}^{n} p(x_i/\pi_i)$$

joint pdf factorized form modeled by the Bayesian network

$$p(A, B, X, C, D) = p(A)p(B)p(X/A, B)p(C/X)p(D/X)$$