EE7403 LEC 10 Discriminant Functions and Classifiers.

1. General classification process

MAP: Decide $w_k = \arg\max_{w_i} [p(w_i|x)] = \arg\min_{w_i} [p(e_i|x)]$

discriminant functions:

$$g_i(x) = \ln p(x|w_i) + \ln p(w_i) \quad [-\ln p(x) \text{ is fixed}].$$

2. Discriminant function for multivariate Gaussian PDF

If: class conditional PDF is multi-variate Gaussian of

$$p(x|w_i) = N(\mu_i, \Sigma_i)$$

$$= \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)\right]$$

$$g_i(x) = \ln p(x|w_i) + \ln p(w_i)$$

hyperquadrics

$$= -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) + \ln p(w_i) - \frac{1}{2}\ln|\Sigma_i| \cancel{- \frac{d}{2}\ln 2\pi}$$

$$= -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) + b_i$$

$$= -\frac{1}{2} \underbrace{d_\Sigma(x, \mu_i)}_{\text{Mahalanobis distance}} + b_i$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n} (x_k - \hat{\mu})(x_k - \hat{\mu})^T$$

In 1-D case. $\frac{(x-\mu_i)^2}{\sigma^2}$

Mahalanobis distance: $d_\Sigma(x, \mu_i) = (x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)$

Euclidean distance: $d_{Eu}(x, \mu_i) = (x-\mu_i)^T (x-\mu_i)$ $\qquad (x-\mu_i)^2$

① Special case 1: all classes own the same covariance $\Sigma_i = \Sigma$

☆ $$\boxed{g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) + \ln p(w_i)}$$

$\Downarrow$

+ same prior propability

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) = d_\Sigma(x, \mu_i)$$

distance.

minimum Mahalanobis classifier (for any covariance Σ, $\Sigma = \Sigma^T$)

$\Sigma_1 = \Sigma_2 \cdots = \Sigma_c$

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma^{-1}(x-\mu_i) + \ln P(w_i)$$

$$= -\frac{1}{2}x^T\Sigma^{-1}x + \mu_i^T \Sigma^{-1} x - \frac{1}{2}\mu_i^T \Sigma \mu_i + \ln(P(w_i))$$

drop the variable irrelative to i: $g_i(x) = \underbrace{(\mu_i^T \Sigma^{-1})}_{= w_i^T} x \underbrace{\left(-\frac{1}{2}\mu_i^T \Sigma^{-1}\mu_i + \ln(P(w_i))\right)}_{= w_{i0}}$

$= w_i^T x + w_{i0}$ threshold / bias

linear classifier

boundary hyperplane.

the decision boundary: $g_i(x) = g_j(x) \Rightarrow 0 = (\mu_i - \mu_j)^T \Sigma^{-1} x - \frac{1}{2}(\mu_i - \mu_j)^T \Sigma (\mu_i - \mu_j) + \ln\left[\frac{P(w_i)}{P(w_j)}\right]$

the prior probability shift the optimal boundary hyper-plane.

EE7403 LEC10

③ special case 2: all classes own the same diagonal, scalar covariance matrix $\Sigma_i = \sigma^2 I \Rightarrow \Sigma_i^{-1} = \frac{I}{\sigma^2} \quad |\Sigma_i| = \sigma^{2d}$

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma^{-1} (x-\mu_i) + \ln P(w_i)$$

$$= -\frac{1}{2\sigma^2}(x-\mu_i)^T (x-\mu_i) + \ln P(w_i)$$

$$= -\frac{1}{2\sigma^2}(x^T x - 2\mu_i^T x + \mu_i^T \mu_i) + \ln P(w_i)$$

$$\Rightarrow g_i = \mu_i^T x - \frac{\mu_i^T \mu_i}{2} + \sigma^2 \ln P(w_i)$$

$$= w_i^T x + w_{i0}$$

classification boundary $g_i = g_j \Rightarrow (w_i - w_j)^T x + w_{i0} - w_{j0} = 0$

$$0 = (\mu_i - \mu_j)^T x \quad \frac{1}{2}(\mu_i - \mu_j)^T (\mu_i - \mu_j) + \sigma^2 \ln\left[\frac{P(w_i)}{P(w_j)}\right]$$

define a hyperplane orthogonal to the line linking the 2 means

which means: no feature has more influence.
no class is favored
depending only on mean separation. (for $P(w_i)$ equal)

$$g_i(x) = -\frac{1}{2\sigma^2}(x-\mu_i)^T (x-\mu_i) + \ln p(w_i)$$

for $P(w_i)$ are the same

$g_i(x) = -(x-\mu_i)^T (x-\mu_i) = -d_{Eu}(x, \mu_i) = \|x-\mu_i\| \Rightarrow$ minimum Euclidean distance classifier.

EE 7403 LEC 10

3. Classifier by Sparse Representation.

Basic idea. Query image y can be well represented by a linear combination of its training images.

$A_i = [a_{i,1}, \dots, a_{i,j}, \dots a_{i,n_i}]$: training images of ith class

$$y = A_i x_i + e_i$$

$A = [A_1, A_2, \dots, A_i, \dots, A_c]$: training samples of all c subjects.

$$\underline{y = Ax + e} \quad \text{here } x = [x_1; \dots; x_i; \dots; x_c]$$

If $x = [0; \dots; 0; x_i; 0; \dots; 0]$, we can identify the class i of query image y.

Give y and A; find sparse coefficients x

$$\boxed{\min_x \|y - Ax\|_2^2 + \lambda \|x\|_1}$$

when images are sudiveded into small points, some blocks can be unreliable.

Single image → separete (X) get training samples (✓)

$$\underline{y = d + b + s}$$

d: class-specific / identity component

b: non-class-specific / intra-class variantion component

s: sparse noise or corruption

what is its training samples?

Supervised low-rank decomposition (SLR):

$$\min_{D, B, \Gamma, E} \|D\|_* + \lambda \|B\|_* + \tau \|\Gamma\|_F^2 + \eta \|E\|_1$$

$$\text{s.t. } \underline{A = D + B\Gamma + E}$$