Probabilistic Models for Classification

Monday, 18 August 2025 22:25

Static pattern classification:

Example is represented by a feature vector of dimension d

Sequential pattern classification: Text, Audio, Video

Example is represented by a sequence of feature vectors of dimension d

Binary classification : No. of classes = 2

Multi-class classification : No. of classes > 2 (L)

Training examples :

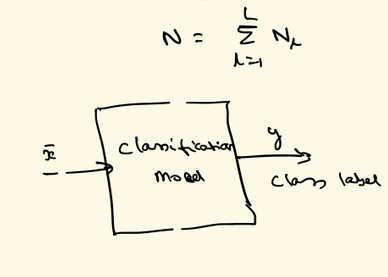
$$D = \{\bar{x}_n, t_n\}_{n=1}^{N}$$

Set of class labels: $\{y_1, y_2, \dots y_L\}$

$t_n \in \{y_1, y_2, \dots y_L\}$

No. of examples of lth class: N_l

$$N = \sum_{l=1}^{L} N_l$$



K Nearest Neighbours classifier :

No model

K = 1 Nearest Neighbour classifier :

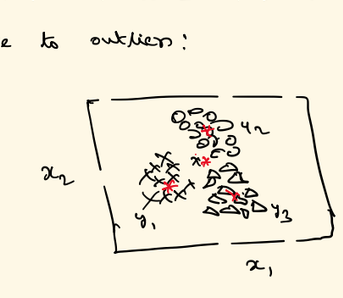
Test example : $\bar{x}$ y ?

$$n^* = \underset{n}{\operatorname{argmin}} \|\bar{x} - \bar{x}_n\|$$

$\|\bar{x} - \bar{x}_{n^*}\| < \|\bar{x} - \bar{x}_n\|$ for $n \neq n^*$

class label for $\bar{x}$ = class label of $\bar{x}_{n^*}$

Sensitive to outliers :



KNN : $K > 1$

Find the K nearest neighbours of $\bar{x}$

L=5 K=10 K << N

N=1000

N_l = 200

$z_n = \|\bar{x} - \bar{x}_n\|$, $n = 1, 2, \dots, N$

Sort $\{z_n\}_{n=1}^{N}$ in increasing order

Top K in the sorted list

K_l : No. of examples among top K belonging to l^{th} class.

$0 \le K_l \le K$

K_1	K_2	K_3	K_4	K_5
1	3	2	1	3

$$l^* = \underset{l}{\operatorname{argmax}}\ K_l$$

class label for $\bar{x}$: y_{l^*}

In case of a tie :

There is more than one class with maximum of K_ls.

$\tilde{z}_l$: Average distance of examples among the K NNs belonging to l^{th} class.

$$l^* = \underset{l}{\operatorname{argmin}}\ \tilde{z}_l$$

(tied classes)

class label of $\bar{x}$ = y_{l^*}

K is a hyperparameter

- Validation method to decide K

Limitation :

computationally intensive during the inference phase

- N distances $\|\bar{x} - \bar{x}_n\|^2$
- Sort N distances $O(N \log N)$

Bayes classifier :

Probabilistic model

Statistical characteristics of data of each of the classes

X and Y are random variables

$P(X, Y)$ $P(X)$ $P(Y)$

Sum rule : Marginalization

$$P(X) = \sum_{Y} P(X, Y)$$

$$P(Y) = \sum_{X} P(X, Y)$$

Product rule :

$P(X,Y) = P(X/Y)\,P(Y)$

$P(X,Y) = P(Y/X)\,P(X)$

Bayes rule :

$$P(Y/X) = \frac{P(X/Y)\,P(Y)}{P(X)}$$

Classification :

$\bar{x}$ Feature vector, Y class label

$$P(Y/\bar{x}) = \frac{P(\bar{x}/Y)\,P(Y)}{P(\bar{x})}$$

Set of L classes : $\{y_1, y_2, \dots y_l, \dots, y_L\}$

$P(Y=y_l/\bar{x})$: Posterior probability for y_l

$p(\bar{x}/Y=y_l)$: class-conditional probability density fn. for y_l

$P(\bar{x})$: Evidence term, Normalization factor

$P(Y=y_l)$: Prior probability for y_l

Adult - child classification

Height H, Weight W (x_1, x_2)

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

y_1 : Adult y_2 : Child

N_1 D_1 N_2 D_2

H: 135-200   H: 30-180

W: 45-100   W: 5-80

$p(\bar{x}/Y=y_1)$

$\bar{x} = \begin{bmatrix} 80 \\ 30 \end{bmatrix}$

$p(\bar{x}/Y=y_2)$ $p(\bar{x}/Y=y_2)$

$\bar{x} = \begin{bmatrix} 20 \\ 10 \end{bmatrix}$ $\bar{x} = \begin{bmatrix} 250 \\ 100 \end{bmatrix}$

$\bar{x} = \begin{bmatrix} 140 \\ 60 \end{bmatrix}$

$$P(Y=y_l/\bar{x}) = \frac{p(\bar{x}/Y=y_l)\,P(Y=y_l)}{p(\bar{x})}$$

$$p(\bar{x}) = \sum_{Y} p(\bar{x}, Y) = \sum_{m=1}^{L} p(\bar{x}, Y=y_m)$$

$$p(\bar{x}) = \sum_{m=1}^{L} p(\bar{x}/Y=y_m)\cdot P(Y=y_m)$$

$$P(Y=y_l/\bar{x}) = \frac{p(\bar{x}/Y=y_l)\,P(Y=y_l)}{\sum_{m=1}^{L} p(\bar{x}/Y=y_m)\,P(Y=y_m)}$$

$$\sum_{l=1}^{L} P(Y=y_l/\bar{x}) = \frac{\sum_{l=1}^{L} p(\bar{x}/Y=y_l)\,P(Y=y_l)}{\underbrace{\sum_{m=1}^{L} p(\bar{x}/Y=y_m)\,P(Y=y_m)}_{\text{Normalizing factor}}} = 1$$

$\bar{x}$ class label for $\bar{x}$

Bayes decision rule:

$$l^* = \underset{l}{\operatorname{argmax}}\ P(Y=y_l/\bar{x})$$

class label for $\bar{x}$: y_{l^*}

$$P(Y=y_l/\bar{x}) = \frac{p(\bar{x}/Y=y_l)\cdot P(Y=y_l)}{p(\bar{x})}$$

Bayes decision rule

$$l^* = \underset{l}{\operatorname{argmax}}\ p(\bar{x}/Y=y_l)\,P(Y=y_l)$$

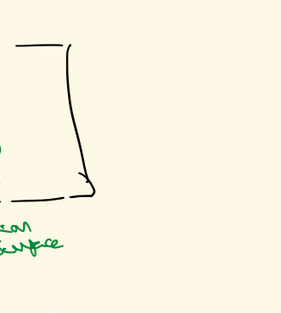
$$l^* = \underset{l}{\operatorname{argmax}}\ \ln p(\bar{x}/Y=y_l) + \ln P(Y=y_l)$$

$$\underset{\uparrow\, x_2}{y} = m\underset{\uparrow\, x_1}{x} + c$$

$$\underbrace{w_2 x_2 + w_1 x_1 + w_0}_{g(\bar{x})} = 0$$

$x_2 + x_1 - 1 = 0$

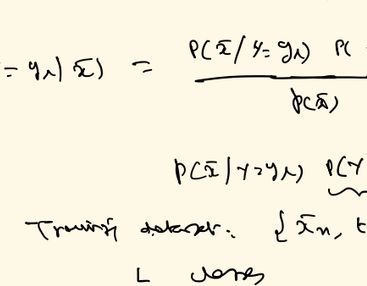
$x_2 = -x_1 + 1$



Discriminant fn.

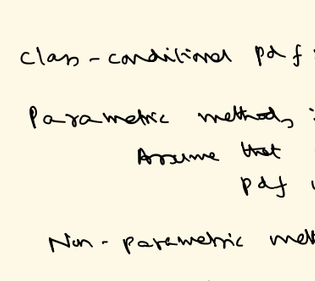
$g(\bar{x}) > 0$ Positive region

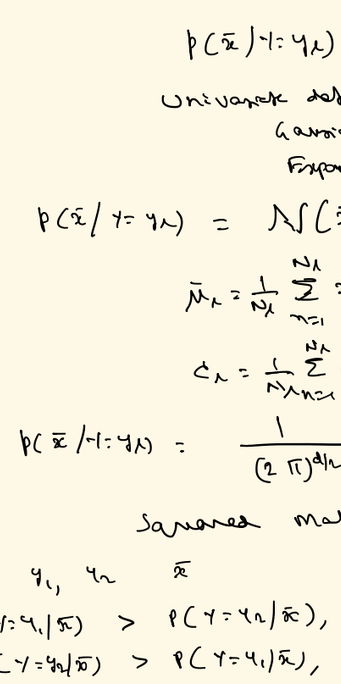
< 0 Negative region



$$\underbrace{x_2^2 + x_1^2 - 1}_{g(\bar{x})} = 0$$

Decision Regions and Decision Surface:





$$P(Y=y_l/\bar{x}) = \frac{p(\bar{x}/Y=y_l)\,P(Y=y_l)}{p(\bar{x})}$$

$p(\bar{x}/Y=y_l)$ $\underset{\sim}{P(Y=y_l)}$

Training dataset : $\{\bar{x}_n, t_n\}_{n=1}^{N}$

L classes

N_l examples for class y_l.

$$N = \sum_{l=1}^{L} N_l$$

$$P(Y=y_l) = \frac{N_l}{N}$$

class-conditional pdf : $p(\underset{\uparrow}{\bar{x}}/Y=y_l)$

Parametric methods :

Assume that class-conditional pdf is of a particular form

Non-parametric methods:

$p(\bar{x}/Y=y_l)$ Gaussian

Univariate data $p(x)$

Gaussian, Cauchy

Exponential, Rayleigh, Maxwell

$$p(\bar{x}/Y=y_l) = \mathcal{N}(\bar{x}/\underset{\uparrow}{\bar{\mu}_l}, \underset{\uparrow}{C_l})$$

$$\bar{\mu}_l = \frac{1}{N_l}\sum_{n=1}^{N_l} \bar{x}_n$$

$$C_l = \frac{1}{N_l}\sum_{n=1}^{N_l} (\bar{x}_n - \bar{\mu}_l)(\bar{x}_n - \bar{\mu}_l)^t$$

$$p(\bar{x}/Y=y_l) = \frac{1}{(2\pi)^{d/2}\,|C_l|^{1/2}}\, e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_l)^t C_l^{-1} (\bar{x}-\bar{\mu}_l)}$$

Squared Mahalanobis distance

$L = 2$ y_1, y_2 $\bar{x}$

If $P(Y=y_1/\bar{x}) > P(Y=y_2/\bar{x})$, assign y_1 to $\bar{x}$

If $P(Y=y_2/\bar{x}) > P(Y=y_1/\bar{x})$, assign y_2 to $\bar{x}$

If $g_1(\bar{x}) > g_2(\bar{x})$, assign y_1 to $\bar{x}$

If $g_2(\bar{x}) > g_1(\bar{x})$, assign y_2 to $\bar{x}$

Equivalent forms of $g_l(\bar{x})$

$$\left\{\begin{aligned} g_l(\bar{x}) &= P(Y=y_l/\bar{x}) \\ g_l(\bar{x}) &= p(\bar{x}/Y=y_l)\,P(Y=y_l) \\ g_l(\bar{x}) &= \ln p(\bar{x}/Y=y_l) + \ln P(Y=y_l) \end{aligned}\right.$$

Decision surface:

$$\underbrace{g_1(\bar{x}) - g_2(\bar{x})}_{g_{12}(\bar{x})} = 0$$

Discriminant function g

surface

If $g_{12}(\bar{x}) > 0$, assign y_1 to $\bar{x}$

If $g_{12}(\bar{x}) < 0$, assign y_2 to $\bar{x}$

$p(\bar{x}/Y=y_1) = \mathcal{N}(\bar{x}/\bar{\mu}_1, C_1)$

$p(\bar{x}/Y=y_2) = \mathcal{N}(\bar{x}/\bar{\mu}_2, C_2)$

$$g_1(\bar{x}) = \ln p(\bar{x}/Y=y_1) + \ln P(Y=y_1)$$

$$p(\bar{x}/Y=y_1) = \frac{1}{(2\pi)^{d/2}\,|C_1|^{1/2}}\, e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_1)^t C_1^{-1}(\bar{x}-\bar{\mu}_1)}$$

$$\ln p(\bar{x}/Y=y_1) = -\frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|C_1| - \frac{1}{2}(\bar{x}-\bar{\mu}_1)^t C_1^{-1}(\bar{x}-\bar{\mu}_1)$$

$$g_1(\bar{x}) = -\frac{1}{2}(\bar{x}^t - \bar{\mu}_1^t) C_1^{-1}(\bar{x}-\bar{\mu}_1) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|C_1| + \ln P(Y=y_1)$$

$$= -\frac{1}{2}\left[\bar{x}^t C_1^{-1}\bar{x} - \bar{x}^t C_1^{-1}\bar{\mu}_1 - \bar{\mu}_1^t C_1^{-1}\bar{x} + \bar{\mu}_1^t C_1^{-1}\bar{\mu}_1\right] - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|C_1| + \ln P(Y=y_1)$$

$$g_1(\bar{x}) = -\frac{1}{2}\left[\bar{x}^t C_1^{-1}\bar{x} - 2\bar{\mu}_1^t C_1^{-1}\bar{x} + \bar{\mu}_1^t C_1^{-1}\bar{\mu}_1\right] - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|C_1| + \ln P(Y=y_1)$$

$$g_2(\bar{x}) = -\frac{1}{2}\left[\bar{x}^t C_2^{-1}\bar{x} - 2\bar{\mu}_2^t C_2^{-1}\bar{x} + \bar{\mu}_2^t C_2^{-1}\bar{\mu}_2\right] - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|C_2| + \ln P(Y=y_2)$$

$$g_{12}(\bar{x}) = g_1(\bar{x}) - g_2(\bar{x})$$

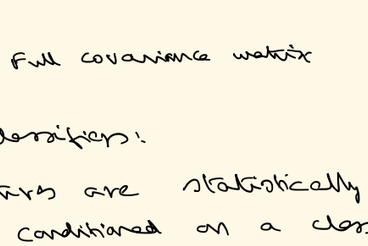
Case 1 : $C_1 = C_2 = C$

$$g_1(\bar{x}) = -\frac{1}{2}\left[\cancel{\bar{x}^t C^{-1}\bar{x}} - 2\bar{\mu}_1^t C^{-1}\bar{x} + \bar{\mu}_1^t C^{-1}\bar{\mu}_1\right] - \cancel{\frac{d}{2}\ln 2\pi} - \cancel{\frac{1}{2}\ln|C|} + \ln P(Y=y_1)$$

$$g_2(\bar{x}) = -\frac{1}{2}\left[\cancel{\bar{x}^t C^{-1}\bar{x}} - 2\bar{\mu}_2^t C^{-1}\bar{x} + \bar{\mu}_2^t C^{-1}\bar{\mu}_2\right] - \cancel{\frac{d}{2}\ln 2\pi} - \cancel{\frac{1}{2}\ln|C|} + \ln P(Y=y_2)$$

$$g_{12}(\bar{x}) = \underbrace{\left[C^{-1}(\bar{\mu}_1 - \bar{\mu}_2)\right]^t}_{\bar{w}}\bar{x} \underbrace{- \frac{1}{2}\bar{\mu}_1^t C^{-1}\bar{\mu}_1 + \frac{1}{2}\bar{\mu}_2^t C^{-1}\bar{\mu}_2 + \ln\frac{P(Y=y_1)}{P(Y=y_2)}}_{w_0} = 0$$

Hyperplane : $\bar{w}^t\bar{x} + w_0 = 0$



Case 2 : $C_1 \neq C_2$

$$g_{12}(\bar{x}) = -\frac{1}{2}\left[\bar{x}^t (C_1^{-1} - C_2^{-1})\bar{x}\right] + (C_1^{-1}\bar{\mu}_1 - C_2^{-1}\bar{\mu}_2)^t\bar{x} - \frac{1}{2}\bar{\mu}_1^t C_1^{-1}\bar{\mu}_1 + \frac{1}{2}\bar{\mu}_2^t C_2^{-1}\bar{\mu}_2 - \frac{1}{2}\ln|C_1| + \frac{1}{2}\ln|C_2| + \ln\frac{P(Y=y_1)}{P(Y=y_2)} = 0$$

$$\underbrace{\bar{x}^t W\bar{x}} + \bar{w}^t\bar{x} + w_0 = 0$$

Hyperquadratic surfaces

Covariance matrices are diagonal :

C_1 and C_2 are diagonal

Naive - Bayes classifier :

Features are uncorrelated for the data of each class.

No. of parameters per class: $\bar{\mu}$, C (d, d)

Full covariance matrix $\frac{d(d+1)}{2}$

Naive - Bayes classifiers :

All features are statistically independent conditioned on a class.

$p(\bar{x}/Y=y_l)$

$\bar{x}$: d-dimensional vector

$P(X, Y) = P(X/Y)\cdot P(Y)$

$= P(X)\,P(Y)$

$p(x_1, x_2) = p(x_1)\,p(x_2)$

$$p(\bar{x}) = \prod_{j=1}^{d} p(x_j) \leftarrow$$

$$p(\bar{x}/Y=y_l) = \prod_{j=1}^{d} p(x_j/Y=y_l) \leftarrow$$

$$p(\bar{x}/y_l) = \frac{1}{(2\pi)^{d/2}\,|C_l|^{1/2}}\, e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_l)^t C_l^{-1}(\bar{x}-\bar{\mu}_l)}$$

$d = 2$

C_l is diagonal

$$|C_l| = \prod_{j=1}^{d} \sigma_{lj}^2$$

$(\bar{x}-\bar{\mu}_l)^t C_l^{-1}(\bar{x}-\bar{\mu}_l)$

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \bar{\mu}_l = \begin{bmatrix} \mu_{l1} \\ \mu_{l2} \end{bmatrix} \quad C_l = \begin{bmatrix} \sigma_{l1}^2 & 0 \\ 0 & \sigma_{l2}^2 \end{bmatrix}$$

$$\begin{bmatrix} (x_1 - \mu_{l1}) & (x_2 - \mu_{l2}) \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_{l1}^2} & 0 \\ 0 & \frac{1}{\sigma_{l2}^2} \end{bmatrix} \begin{bmatrix} (x_1 - \mu_{l1}) \\ (x_2 - \mu_{l2}) \end{bmatrix}$$

$$\frac{(x_1 - \mu_{l1})^2}{\sigma_{l1}^2} + \frac{(x_2 - \mu_{l2})^2}{\sigma_{l2}^2}$$

$$(\bar{x}-\bar{\mu}_l)^t C_l^{-1}(\bar{x}-\bar{\mu}_l) = \sum_{j=1}^{d} \frac{(x_j - \mu_{lj})^2}{\sigma_{lj}^2}$$

$$p(\bar{x}/Y=y_l) = \frac{1}{(2\pi)^{d/2}\prod_{j=1}^{d}\sigma_{lj}}\, e^{-\frac{1}{2}\sum_{j=1}^{d}\frac{(x_j - \mu_{lj})^2}{\sigma_{lj}^2}}$$

$$= \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi}\,\sigma_{lj}}\, e^{-\frac{1}{2}\frac{(x_j - \mu_{lj})^2}{\sigma_{lj}^2}}$$

$$p(\bar{x}/Y=y_l) = \prod_{j=1}^{d} p(x_j/Y=y_l)$$

Decision Points for univariate case :

$L = 2$ y_1, y_2

$$p(x/Y=y_1) = \begin{cases} \theta_1 e^{-\theta_1 x}, & \text{for } x \ge 0 \\ 0, & \text{for } x < 0 \end{cases} \quad \theta_1 > 0$$

Exponential

$$p(x/Y=y_2) = \begin{cases} \theta_2 e^{-\theta_2 x}, & \text{for } x \ge 0 \\ 0, & \text{for } x < 0 \end{cases} \quad \theta_2 > 0$$

Decision point : x^*

If $x > x^*$, assign y_1 to x

If $x < x^*$, assign y_2 to x

$P(Y=y_1) \neq P(Y=y_2)$

$$g_1(x) = \ln p(x/Y=y_1) + \ln P(Y=y_1)$$

$$g_2(x) = \ln p(x/Y=y_2) + \ln P(Y=y_2)$$

$$g_{12}(x) = g_1(x) - g_2(x) = 0$$

$$x^* = \frac{1}{\theta_2 - \theta_1}\left[-\ln\frac{\theta_1}{\theta_2} + \ln\frac{P(Y=y_1)}{P(Y=y_2)}\right]$$

$$g_1(x) = \ln\theta_1 - \theta_1 x + \ln P(Y=y_1)$$

$$g_2(x) = \ln\theta_2 - \theta_2 x + \ln P(Y=y_2)$$

$$g_{12}(x) = (-\theta_1 + \theta_2)x + \ln\frac{\theta_1}{\theta_2} + \ln\frac{P(Y=y_1)}{P(Y=y_2)} = 0$$

$$x^* = \frac{1}{\theta_1 - \theta_2}\left[\ln\frac{\theta_1}{\theta_2} + \ln\frac{P(Y=y_1)}{P(Y=y_2)}\right]$$

Discrete features :

Binary features x_j is 0 or 1

Bernoulli distribution :

$p_j = P(x_j = 1/Y=y_1)$ $p(x_j/Y=y_1) = p_j^{x_j}(1-p_j)^{1-x_j}$

$q_j = P(x_j = 1/Y=y_2)$ $p(x_j/Y=y_2) = q_j^{x_j}(1-q_j)^{1-x_j}$

x_j's are statistically independent

$$p(\bar{x}/Y=y_l) = \prod_{j=1}^{d} p(x_j/Y=y_l)$$

$g_{12}(\bar{x}) = 0$

$$g_1(\bar{x}) = \ln p(\bar{x}/Y=y_1) + \ln P(Y=y_1)$$

$$= \ln\prod_{j=1}^{d} p(x_j/Y=y_1) + \ln P(Y=y_1)$$

$$= \sum_{j=1}^{d} \ln p(x_j/Y=y_1) + \ln P(Y=y_1)$$

$$= \sum_{j=1}^{d} \ln\left[p_j^{x_j}(1-p_j)^{(1-x_j)}\right] + \ln P(Y=y_1)$$

$$g_1(\bar{x}) = \sum_{j=1}^{d} \underline{x_j \ln p_j} + (1-x_j)\ln(1-p_j) + \ln P(Y=y_1)$$

$$g_2(\bar{x}) = \sum_{j=1}^{d} \underline{x_j \ln q_j} + (1-x_j)\ln(1-q_j) + \ln P(Y=y_2)$$

$$g_{12}(\bar{x}) = g_1(\bar{x}) - g_2(\bar{x})$$

$$\sum_{j=1}^{d} x_j \ln\frac{p_j}{q_j} + (1-x_j)\ln\frac{(1-p_j)}{(1-q_j)} + \ln\frac{P(Y=y_1)}{P(Y=y_2)} = 0$$

$$\sum_{j=1}^{d} \underbrace{\ln\frac{p_j(1-q_j)}{q_j(1-p_j)}}_{w_j} x_j + \underbrace{\sum_{j=1}^{d}\ln\frac{1-p_j}{1-q_j} + \ln\frac{P(Y=y_1)}{P(Y=y_2)}}_{w_0} = 0$$

$$\sum_{j=1}^{d} w_j x_j + w_0 = 0$$

Linear surface

$p_j = q_j \rightarrow w_j = 0$

$p_j > q_j$, $1 - p_j < 1 - q_j$ $w_j > 0$

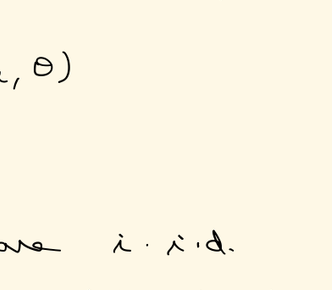
$p_j < q_j$ $w_j < 0$

Multi-valued discrete features statistically independent : Decision surface is linear

Multi-class problem : $L > 2$

$$l^* = \underset{l}{\operatorname{argmax}}\ g_l(\bar{x})$$

$g_i(\bar{x}) = g_j(\bar{x})$ such that $g_j(\bar{x}) > g_k(\bar{x})$, $k \neq j, i$



Parameter Estimation:

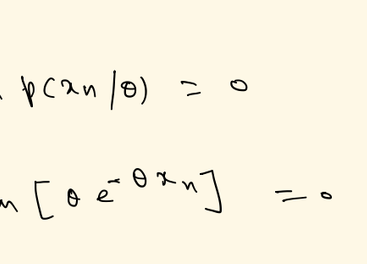
Maximum Likelihood (ML) method

$p(\bar{x}/Y=y_l)$: Likelihood function for class y_l

$$p(x/Y=y_l) = \begin{cases} \theta e^{-\theta x}, & x \ge 0 \\ 0, & x < 0 \end{cases} \quad \theta > 0$$

Training dataset of class y_l : D_l

$D_l = \{x_n\}_{n=1}^{N_l}$



$p(x_n/Y=y_l, \theta)$

Total likelihood for D_l

$\mathcal{L}$

Training examples are i.i.d.

identically and independently distributed.

$\mathcal{L} = p(D_l)$

$$\mathcal{L} = \prod_{n=1}^{N_l} p(x_n/Y=y_l, \theta)$$

Total log likelihood

$$\ln\mathcal{L} = \sum_{n=1}^{N_l} \ln p(x_n/\theta)$$

Find θ for which $\ln\mathcal{L}$ is maximum

$$\frac{d\ln\mathcal{L}}{d\theta} = 0$$

$$\frac{d}{d\theta}\sum_{n=1}^{N_l} \ln p(x_n/\theta) = 0$$

$$\frac{d}{d\theta}\sum_{n=1}^{N_l} \ln\left[\theta e^{-\theta x_n}\right] = 0$$

$$\frac{d}{d\theta}\left[\sum_{n=1}^{N_l} \ln\theta - \theta x_n\right] = 0$$

$$\underbrace{\sum_{n=1}^{N_l}\cdot\frac{1}{\theta}} - \sum_{n=1}^{N_l} x_n = 0$$

$$\frac{N_l}{\theta} = \sum_{n=1}^{N_l} x_n$$

$$\theta_l = \frac{N_l}{\sum_{n=1}^{N_l} x_n}$$

Rayleigh distribution :

$$p(x/\theta) = \begin{cases} 2\theta x e^{-\theta x^2}, & x \ge 0 \\ 0, & x < 0 \end{cases} \quad \theta > 0$$

θ^*

Univariate Gaussian distribution :

$$p(x/\theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}}\, e^{-\frac{(x-\theta_1)^2}{2\theta_2}}$$

$$\theta_1 = \mu = \frac{1}{N}\sum_{n=1}^{N} x_n$$

$$\theta_2 = \sigma^2 = \frac{1}{N}\sum_{n=1}^{N} (x_n - \mu)^2$$

$$\ln\mathcal{L} = \sum_{n=1}^{N} \ln p(x_n/\theta_1, \theta_2)$$

$\frac{\partial\ln\mathcal{L}}{\partial\theta_1} = 0$ θ_1

$\frac{\partial\ln\mathcal{L}}{\partial\theta_2} = 0$, θ_2

Multi-variate Gaussian Distribution:

$$p(\bar{x}/y_l) = \mathcal{N}(\bar{x}/\underset{\uparrow\,\bar{\theta}}{\bar{\mu}_l}, \underset{M}{C_l})$$

Total loglikelihood :

$$\ln\mathcal{L} = \sum_{n=1}^{N} \ln p(\bar{x}_n/\bar{\theta}, M)$$

$$\frac{\partial\ln\mathcal{L}}{\partial\bar{\theta}} = \underset{d\times 1}{\bar{0}}$$

$$\frac{\partial\ln\mathcal{L}}{\partial M} = \underset{d\times d}{0}$$

$$p(\bar{x}_n/\bar{\theta}, M) = \mathcal{N}(\bar{x}_n/\bar{\theta}, M) = \frac{1}{(2\pi)^{d/2}\,|M|^{1/2}}\, e^{-\frac{1}{2}(\bar{x}_n - \bar{\theta})^t M^{-1}(\bar{x}_n - \bar{\theta})}$$

$$\ln p(\bar{x}_n/\bar{\theta}, M) = -\frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|M| - \frac{1}{2}(\bar{x}_n - \bar{\theta})^t M^{-1}(\bar{x}_n - \bar{\theta})$$

$$\frac{\partial\ln p(\bar{x}_n/\bar{\theta}, M)}{\partial\bar{\theta}} = -\frac{1}{2}\frac{\partial}{\partial\bar{\theta}}\left[\bar{x}_n^t M^{-1}\bar{x}_n - 2\bar{\theta}^t \underbrace{M^{-1}\bar{x}_n} + \underbrace{\bar{\theta}^t M^{-1}\bar{\theta}}\right]$$

$$= -\frac{1}{2}\left[-2M^{-1}\bar{x}_n + 2M^{-1}\bar{\theta}\right]$$

$$\frac{\partial\ln\mathcal{L}}{\partial\bar{\theta}} = \sum_{n=1}^{N} M^{-1}\bar{x}_n - \sum_{n=1}^{N} M^{-1}\bar{\theta} = \bar{0}$$

$$M^{-1}\left[\sum_{n=1}^{N}\bar{x}_n - \sum_{n=1}^{N}\bar{\theta}\right] = \bar{0}$$

$$\underbrace{MM^{-1}}_{I}\left[\quad\right] = \underbrace{M\bar{0}}_{\bar{0}}$$

$$\sum_{n=1}^{N}\bar{x}_n - N\bar{\theta} = \bar{0}$$

$$\bar{\theta} = \frac{1}{N}\sum_{n=1}^{N}\bar{x}_n = \bar{\mu}$$

$$\frac{\partial\ln p(\bar{x}_n/\bar{\theta}, M)}{\partial M} = \frac{\partial}{\partial M}\left[-\frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|M| - \frac{1}{2}(\bar{x}_n - \bar{\mu})^t M^{-1}(\bar{x}_n - \bar{\mu})\right]$$

$$\frac{\partial\ln|M|}{\partial M} = M^{-1}$$

$$\frac{\partial(\bar{x}_n - \bar{\mu})^t M^{-1}(\bar{x}_n - \bar{\mu})}{\partial M} =$$

$$\frac{\partial}{\partial M}(\bar{a}^t M^{-1}\bar{a}) = -M^{-1}\bar{a}\bar{a}^t M^{-1}$$

$$\frac{\partial\ln p(\bar{x}_n/\bar{\theta}, M)}{\partial M} = \left[-\frac{1}{2}M^{-1} - \frac{1}{2}\left(-M^{-1}(\bar{x}_n - \bar{\mu})(\bar{x}_n - \bar{\mu})^t M^{-1}\right)\right]$$

$$= -\frac{1}{2}M^{-1} + \frac{1}{2}M^{-1}(\bar{x}_n - \bar{\mu})(\bar{x}_n - \bar{\mu})^t M^{-1}$$

$$\frac{\partial\ln\mathcal{L}}{\partial M} = -\cancel{\frac{1}{2}}\sum_{n=1}^{N} M^{-1} + \cancel{\frac{1}{2}} M^{-1}\sum_{n=1}^{N}(\bar{x}_n - \bar{\mu})(\bar{x}_n - \bar{\mu})^t M^{-1} = 0$$

$$NM^{-1} = M^{-1}\sum_{n=1}^{N}(\bar{x}_n - \bar{\mu})(\bar{x}_n - \bar{\mu})^t M^{-1}$$

$$N\cdot\underbrace{MM^{-1}}_{I}M = \underbrace{MM^{-1}}_{I}\sum_{n=1}^{N}(\bar{x}_n - \bar{\mu})(\bar{x}_n - \bar{\mu})^t\underbrace{M^{-1}M}_{I}$$

$$M = \frac{1}{N}\sum_{n=1}^{N}(\bar{x}_n - \bar{\mu})(\bar{x}_n - \bar{\mu})^t = C$$

$$p(\bar{x}/Y=y_l) = \mathcal{N}(\bar{x}/\bar{\mu}_l, C_l)$$

Gaussianity or normality test

Gaussian distribution: Unimodal distribution (peak)

Multi-modal distribution : Two or more peaks

Gaussian mixture model (GMM)

$$p(\bar{x}/Y=y_l) = \sum_{q=1}^{Q} w_{lq}\underbrace{\mathcal{N}(\bar{x}/\bar{\mu}_{lq}, C_{lq})}$$

Set of parameters : $\{w_{lq}, \bar{\mu}_{lq}, C_{lq}\}_{q=1}^{Q}$ (1, d, $\frac{d(d+1)}{2}$)

No. of parameters : $Q\left(1 + d + \frac{d(d+1)}{2}\right)$

Overlapping Gaussians