d

K

$$\mathbf{x} = [x_1, \cdots, x_d]^T$$

$x_i \ (i = 1, \cdots, d)$

$$\{C_1, \cdots, C_K\}$$

$\mathbf{x} \in C_k$

N

$$\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]$$

$$\mathbf{y} = [y_1, \cdots, y_N]^T$$

$y_k \in \{1, \cdots, K\}$

$\mathbf{x}_k \in C_{y_k}$

$$K = 2$$

C_+

C_-

$\mathbf{x}_j \ (j = 1, \cdots, N)$

$$y_j = 1$$

$\mathbf{x}_j \in C+$

$$y_j = -1$$

$\mathbf{x}_j \in C_-$

N_k

$$\{\mathbf{x}_1^{(k)}, \cdots, \mathbf{x}_{N_k}^{(k)}\}$$

$$C_k, \quad (k = 1, \cdots, K)$$

$$N = \sum_{k=1}^{K} N_k$$

$$P_k = N_k/N$$

C_k

$$\mathbf{x}_1, \cdots, \mathbf{x}_N$$

$C_1, \cdots, C_K$

$$\{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$$

$$\{y_1, \cdots, y_N\}$$

$\mathbf{x}_n$

$$y_n = k \in \{1, \cdots, K\}$$

k

$$k = 1$$

$$\{\mathbf{x}_1, \cdots, \mathbf{x}_{N_k}\}$$

$k = 1, \cdots, K$

$$\mathbf{m}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} \mathbf{x}_n, \qquad \mathbf{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} (\mathbf{x}_n - \mathbf{m})^T (\mathbf{x}_n - \mathbf{m})$$

$d(\mathbf{x},\, C_k)$

$$d(\mathbf{x}, C_k) = min\{d(\mathbf{x}, C_i) \;\; i = 1, \cdots, C\}$$

$$\mathbf{x} \in C_k$$

$d_E(\mathbf{x}, \mathbf{m}_k)$

$$\mathbf{m}_k$$

$$\Sigma_k$$

$$x = 1$$

$$C_1 \sim \mathcal{N}(5, 1.2^2), \qquad C_2 \sim \mathcal{N}(-5, 3^2)$$

$$m_1 = 5$$

$$m_2 = -5$$

$$d_E(x, m_1) = 4 \; < \; d_E(x, m_2) = 6$$

m_1

m_2

C_1

x

C_2

$\sigma_1 = 1.2$

$$\sigma_2 = 3$$

$d_E(x, m_k)$

σ_k

$$d(x, C_k) = (x - m_k)^2/\sigma_k^2$$

$$d(x_1,\, m_1) = 11.11 \;>\; d(x_1,\, m_2) = 4.0$$

$$d_M(\mathbf{x}, C_k) = (\mathbf{x} - \mathbf{m}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \mathbf{m}_k)$$

$$x_1 = 1$$

$$x_2 = 3$$

$$C_1 \sim \mathcal{N}(0, 1.2^2), \qquad C_2 \sim \mathcal{N}(0, 3^2)$$

$$m_1 = m_2 = 0$$

$$|x_i - m_1|^2 = |x_i - m_2|^2$$

x_1

x_2

$$\sigma_2 > \sigma_2$$

$$\begin{cases} d_M(x_1,\, C_1) = \frac{(x_1 - m_1)^2}{\sigma_1^2} = \frac{1}{1.2^2} = 0.69, & d_M(x_1,\, C_2) = \frac{(x_1 - m_1)^2}{\sigma_2^2} = \frac{1}{9} = 0.11 \\ d_M(x_2,\, C_1) = \frac{(x_1 - m_1)^2}{\sigma_1^2} = \frac{3}{1.2^2} = 6.25, & d_M(x_1,\, C_2) = \frac{(x_1 - m_1)^2}{\sigma_2^2} = \frac{3^2}{9} = 1 \end{cases}$$

$$P(C_k|\mathbf{x}) = \frac{p(\mathbf{x}, C_k)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|C_k)P(C_k)}{p(\mathbf{x})}$$

$$P(C_k)$$

P_k

$$P_k = \frac{N_k}{N} = \frac{N_k}{\sum_{l=1}^{K} N_l}, \qquad (k = 1, \cdots, K)$$

$$p(\mathbf{x}|C_k) = L(C_k|\mathbf{x})$$

$$p(\mathbf{x}|C_k) = N(\mathbf{x}, \mathbf{m}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{d/2} \left|\boldsymbol{\Sigma}_k\right|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \mathbf{m}_k)\right]$$

$$p(\mathbf{x}, C_k)$$

$$p(\mathbf{x}, C_k) = p(\mathbf{x}|C_k)P(C_k) = P(C_k|\mathbf{x})p(\mathbf{x})$$

$p(\mathbf{x})$

$$p(\mathbf{x}|C_k)$$

$$p(\mathbf{x}) = \sum_{k=1}^{K} p(\mathbf{x}, C_k) = \sum_{k=1}^{K} P_k \, p(\mathbf{x}|C_k) = \sum_{k=1}^{K} P_k \, \frac{1}{(2\pi)^{d/2} \left|\boldsymbol{\Sigma}_k\right|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \mathbf{m}_k)\right]$$

$$P(C_k|\mathbf{x})$$

$$P(C_k|\mathbf{x}) = \max_l \{P(C_l|\mathbf{x}),\ (l = 1, \cdots, K)\},$$

$$p(\mathbf{x}|C_k)\,P_k$$

R_1

R_2

$$P(\mathbf{x} \in C_i \cap \mathbf{x} \in R_j)$$

$$\mathbf{x} \in C_i$$

R_j

$$P(error)$$

$=$

$$P((\mathbf{x} \in R_2) \cap (\mathbf{x} \in C_1)) + P((\mathbf{x} \in R_1) \cap (\mathbf{x} \in C_2))$$

$$P(\mathbf{x} \in R_2/C_1)\, P_1 + P(\mathbf{x} \in R_1/C_2)\, P_2$$

$$P_1 \int_{R_2} p(\mathbf{x}/C_1) d\mathbf{x} + P_2 \int_{R_1} p(\mathbf{x}/C_2) d\mathbf{x}$$

$$p(\mathbf{x}|C_i)P_i = p(\mathbf{x}|C_j)P_j$$

C_i

$$C_j$$

$$p(\mathbf{x}|C_k) = \mathcal{N}(\mathbf{x}, \mathbf{m}_k, \mathbf{\Sigma}_k)$$

$$\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$$

$$\mathbf{m}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbf{x}_i, \qquad \boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} (\mathbf{x}_i - \mathbf{m}_k)(\mathbf{x}_i - \mathbf{m}_k)^T, \qquad (\mathbf{x}_i \in C_k)$$

$$\log P(C_k|\mathbf{x})$$

$$\log\left[p(\mathbf{x}|C_k)P_k/p(\mathbf{x})\right] = \log p(\mathbf{x}|C_k) + \log P_k - \log p(\mathbf{x})$$

$$-\frac{1}{2}(\mathbf{x}-\mathbf{m}_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}-\mathbf{m}_k)-\frac{N}{2}\log(2\pi)-\frac{1}{2}\log|\boldsymbol{\Sigma}_k|+\log P_k-\log p(\mathbf{x})$$

$$\log p(\mathbf{x})$$

$N \log(2\pi)/2$

$$D_k(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\mathbf{m}_k)^T \mathbf{\Sigma}_k^{-1}(\mathbf{x}-\mathbf{m}_k) - \frac{1}{2}\log|\mathbf{\Sigma}_k| + \log P_k$$

$$D_l(\mathbf{x}) = \max\{D_k(\mathbf{x}),\ (k = 1, \cdots, K)\},$$

$$\mathbf{x} \in C_l$$

$$d_k(\mathbf{x}) = (\mathbf{x} - \mathbf{m}_k)^T \mathbf{\Sigma}_k^{-1} (\mathbf{x} - \mathbf{m}_k) + \log |\mathbf{\Sigma}_k| - 2 \log P_k$$

$$D_i(\mathbf{x}) = D_j(\mathbf{x})$$

$$-\frac{1}{2}(\mathbf{x}-\mathbf{m}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\mathbf{m}_i)-\frac{1}{2}\ \log\ |\boldsymbol{\Sigma}_i|+\log\ P_i$$

$$-\frac{1}{2}(\mathbf{x}-\mathbf{m}_j)^T \mathbf{\Sigma}_j^{-1}(\mathbf{x}-\mathbf{m}_j) - \frac{1}{2}\ \log\ |\mathbf{\Sigma}_j| + \log\ P_j$$

$$\mathbf{x}^T \mathbf{W} \mathbf{x} + \mathbf{w}^T \mathbf{x} + w = 0$$

W

$$-\frac{1}{2}(\mathbf{\Sigma}_i^{-1} - \mathbf{\Sigma}_j^{-1})$$

W

$$\boldsymbol{\Sigma}_i^{-1}\mathbf{m}_i - \boldsymbol{\Sigma}_j^{-1}\mathbf{m}_j$$

w

$$-\frac{1}{2}(\mathbf{m}_i^T \boldsymbol{\Sigma}_i^{-1}\mathbf{m}_i - \mathbf{m}_j^T \boldsymbol{\Sigma}_j^{-1}\mathbf{m}_j) - \frac{1}{2}\log\ \frac{|\boldsymbol{\Sigma}_i|}{|\boldsymbol{\Sigma}_j|} + \log\ \frac{P_i}{P_j}$$

$$\mathbf{x}^T \mathbf{W} \mathbf{x} + \mathbf{w}^T \mathbf{x} + w \begin{cases} > 0 \\ < 0 \end{cases},$$

$$\mathbf{x} \in \left\{ \begin{array}{l} C_i \\ C_j \end{array} \right.$$

$$P_i = P_j \quad (i, j = 1, \cdots, K)$$

$$D_i(\mathbf{x})$$

$$D_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mathbf{m}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \mathbf{m}_i) - \frac{1}{2} \ \log \ |\boldsymbol{\Sigma}_i|$$

$$d(\mathbf{x},\, C_i) = (\mathbf{x} - \mathbf{m}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \mathbf{m}_i) + \log\, |\boldsymbol{\Sigma}_i|$$

$$\begin{cases} d(x_1, C_1) = 0.69 + \log(1.2^2) = 1.06, & d_M(x_1, C_2) = 0.11 + \log(3^2) = 2.31 \\ d_M(x_2, C_1) = 6.25 + \log(1.2^2) = 6.61, & d_M(x_1, C_2) = 1 + \log(3^2) = 3.20 \end{cases}$$

$$\log(\sigma^2)$$

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$$

$$D_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mathbf{m}_i)^T \mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{m}_i) + \log\ P_i$$

P_i

$$-\frac{1}{2}(\mathbf{x}-\mathbf{m}_i)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{m}_i)+\log\, P_i=-\frac{1}{2}(\mathbf{x}-\mathbf{m}_j)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{m}_j)+\log\, P_j$$

$$\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x}$$

$$\mathbf{w}^T \mathbf{x} + w = 0$$

$$\mathbf{w} = \boldsymbol{\Sigma}^{-1}(\mathbf{m}_i - \mathbf{m}_j)$$

$$w = -\frac{1}{2}(\mathbf{m}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{m}_i - \mathbf{m}_j^T \boldsymbol{\Sigma}^{-1} \mathbf{m}_j) + \log \frac{P_i}{P_j}$$

$\mathbf{m}_i$

$$\mathbf{m}_j$$

$$\boldsymbol{\Sigma}^{-1}(\mathbf{m}_i - \mathbf{m}_j)$$

$$\mathbf{m}_i - \mathbf{m}_j$$

$$\mathbf{\Sigma}^{-1}$$

$$\boldsymbol{\Sigma}_i = \sigma^2 \, \mathbf{I} = diag[\sigma^2, \cdots, \sigma^2]$$

$$|\boldsymbol{\Sigma}_i| = \sigma^{2d}$$

$$D_i(\mathbf{x}) = -\frac{||\mathbf{x} - \mathbf{m}_i||^2}{2\sigma^2} + \log\ P_i$$

$$\log |\mathbf{\Sigma}_i|$$

$$\frac{||\mathbf{x}-\mathbf{m}_i||^2}{2\sigma^2}-\log\ P_i=\frac{||\mathbf{x}-\mathbf{m}_j||^2}{2\sigma^2}-\log\ P_j$$

$$\mathbf{w} = \mathbf{m}_i - \mathbf{m}_j, \qquad w = -(\mathbf{m}_i^T \mathbf{m}_i - \mathbf{m}_j^T \mathbf{m}_j) + 2\sigma^2 \ \log \ \frac{P_i}{P_j}$$

$$D_i(\mathbf{x}) = -(\mathbf{x} - \mathbf{m}_i)^T(\mathbf{x} - \mathbf{m}_i) = -||\mathbf{x} - \mathbf{m}_i||^2$$

$$||\mathbf{x} - \mathbf{m}_i||^2$$

$$9/200 = 4.5\%$$

$$\begin{bmatrix} 92 & 8 \\ 1 & 99 \end{bmatrix}$$

$$61/400 = 15.25\%$$

$$\begin{bmatrix} 175 & 25 \\ 36 & 164 \end{bmatrix}$$

$$49/600 = 8.17\%$$

$$\begin{bmatrix} 196 & 3 & 1 \\ 0 & 191 & 9 \\ 3 & 33 & 164 \end{bmatrix}$$

$N = 256$

$$\begin{bmatrix} 215 & 0 & 0 & 0 & 0 & 2 & 4 & 0 & 3 & 0 \\ 0 & 216 & 0 & 0 & 1 & 0 & 1 & 3 & 1 & 2 \\ 2 & 0 & 219 & 0 & 0 & 0 & 0 & 1 & 2 & 0 \\ 1 & 0 & 0 & 212 & 0 & 1 & 0 & 3 & 4 & 3 \\ 1 & 0 & 1 & 0 & 213 & 0 & 0 & 0 & 4 & 5 \\ 1 & 0 & 2 & 1 & 0 & 211 & 0 & 0 & 9 & 0 \\ 2 & 6 & 0 & 0 & 0 & 0 & 213 & 0 & 3 & 0 \\ 2 & 0 & 3 & 0 & 0 & 0 & 1 & 205 & 3 & 10 \\ 1 & 0 & 2 & 4 & 0 & 8 & 1 & 4 & 200 & 4 \\ 0 & 0 & 1 & 1 & 8 & 0 & 0 & 2 & 2 & 210 \end{bmatrix}$$

$$\{(\mathbf{x}_i,\, y_i) \mid i = 1, \cdots, N\}$$

$\mathbf{x}_i$

$$y_i \in \{-1, +1\}$$

$$h_t(\mathbf{x}_n) \in \{-1, +1\}$$

50%

$$h_t(\mathbf{x}_n) = y_n = \pm 1$$

$$y_n \, h_t(\mathbf{x}_n) = 1$$

$$h_t(\mathbf{x}_n) = -y_n = \pm 1$$

$$y_n \, h_t(\mathbf{x}_n) = -1$$

$$\varepsilon_t = \frac{\sum_{y_n\ h_t(\mathbf{x}_n)=-1} w_t(n)}{\sum_{n=1}^{N} w_t(n)}$$

$$w_t(n)$$

$$1 - \varepsilon_t = 1 - \frac{\sum_{y_n\, h_t(\mathbf{x}_n) = -1} w_t(n)}{\sum_{n=1}^{N} w_t(n)} = \frac{\sum_{y_n\, h_t(\mathbf{x}_n) = 1} w_t(n)}{\sum_{n=1}^{N} w_t(n)}$$

$t = 0$

$$w_0(1) = \cdots = w_0(N) = 1$$

$$\varepsilon_0 = \frac{1}{N} \sum_{y_n \, h_0(\mathbf{x}_n) = -1} w_0(n) = \frac{\text{number of misclassified samples}}{\text{total number of samples}}$$

$$h_0$$

$$\varepsilon_t < 1/2$$

$t > 0$

$$H_t(\mathbf{x}_n)$$

$$h_1(\mathbf{x}_n), \cdots, h_t(\mathbf{x}_n)$$

$$n = 1, \cdots, N$$

$$F_t(\mathbf{x}_n)$$

$$\alpha_t h_t(\mathbf{x}_n) + F_{t-1}(\mathbf{x}_n) = \alpha_t h_t(\mathbf{x}_n) + \alpha_{t-1} h_{t-1}(\mathbf{x}_n) + F_{t-2}(\mathbf{x}_n)$$

$$\cdots = \sum_{i=1}^{t} \alpha_i h_i(\mathbf{x}_n)$$

α_i

$$h_i(\mathbf{x}_n)$$

$$F_t(\mathbf{x})$$

$$H_t(\mathbf{x}_n) = sign[F_t(\mathbf{x}_n)] = \begin{cases} +1 & F_t(\mathbf{x}_n) > 0 \\ -1 & F_t(\mathbf{x}_n) < 0 \end{cases}$$

$$|F_t(\mathbf{x}_n)|$$

$(t > 0)$

$$w_t(n) = e^{-y_n F_{t-1}(\mathbf{x}_n)} \begin{cases} > 1 & \text{if } y_n\, F_{t-1}(\mathbf{x}_n) < 0 \text{ (misclassification)} \\ < 1 & \text{if } y_n\, F_{t-1}(\mathbf{x}_n) > 0 \text{ (correct classification)} \end{cases}$$

$$w_{t-1}(n)$$

$$w_t(n)$$

$$e^{-y_n F_{t-1}(\mathbf{x}_n)} = e^{-y_n [\alpha_{t-1} h_{t-1}(\mathbf{x}_n) + F_{t-2}(\mathbf{x}_n)]}$$

$$e^{-\alpha_{t-1}\, y_n h_{t-1}(\mathbf{x}_n)}\; e^{-y_n\, F_{t-2}(\mathbf{x}_n)} = e^{-\alpha_{t-1}\, y_n h_{t-1}(\mathbf{x}_n)}\; w_{t-1}(n)$$

$$w_1(k) = e^{-\alpha_0 \, y_n h_0(\mathbf{x}_n)} w_0(n)$$

$$E_{t+1} = \sum_{n=1}^{N} w_{t+1}(n) = \sum_{n=1}^{N} e^{-y_n \; F_t(\mathbf{x}_n)} = \sum_{n=1}^{N} e^{-\alpha_t \; y_n \; h_t(\mathbf{x}_n)} \; w_t(n) \qquad (t > 1)$$

α_t

E_{t+1}

h_t

$$E_{t+1} = \sum_{n=1}^{N} w_t(n) e^{-\alpha_t \, y_n \, h_t(\mathbf{x}_n)} = \sum_{y_n h_t(\mathbf{x}_n) = -1} w_t(n) e^{\alpha_t} + \sum_{y_n h_t(\mathbf{x}_n) = 1} w_t(n) e^{-\alpha_t}$$

$$\frac{dE_{t+1}}{d\alpha_t} = \sum_{y_n h_t(\mathbf{x}_n)=-1} w_t(n)e^{\alpha_t} - \sum_{y_n h_t(\mathbf{x}_n)=1} w_t(n)e^{-\alpha_t} = 0$$

$$\alpha_t = \frac{1}{2} \ln \left(\frac{\sum_{y_n h_t(\mathbf{x}_n)=1} w_t(n)}{\sum_{y_n h_t(\mathbf{x}_n)=-1} w_t(n)} \right) = \ln \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}} > 0$$

$$1 - \varepsilon_t > 1/2$$

$$e^{\alpha_t} = \sqrt{\frac{1 - \varepsilon_t}{\varepsilon_t}} > 1, \quad e^{-\alpha_t} = \sqrt{\frac{\varepsilon_t}{1 - \varepsilon_t}} < 1$$

ε_t

H_t

$h_1, \cdots, h_t$

t

$t + 1$

$$w_{t+1}(n) = w_t(n)\, e^{-\alpha_t\, y_n h_t(\mathbf{x}_n)} = \begin{cases} w_t(n)\sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}} < w_t(n) & \text{if } y_n h_t(\mathbf{x}_n) = 1 \\ w_t(n)\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}} > w_t(n) & \text{if } y_n h_t(\mathbf{x}_n) = -1 \end{cases}$$

$$y_n h_t(\mathbf{x}_n) = 1$$

$$w_{t+1}(n) < w_t(n)$$

$$y_n h_t(\mathbf{x}_n) = -1$$

$$w_{t+1}(k) > w_t(k)$$

$$h_{t+1}$$

$$\frac{E_{t+1}}{E_t}$$

$$\frac{\sum_{n=1}^{N} w_{t+1}(n)}{\sum_{n=1}^{N} w_t(n)} = \frac{\sum_{n=1}^{N} w_t(n) e^{-\alpha_t \, y_n \, h_t(\mathbf{x}_n)}}{\sum_{n=1}^{N} w_t(k)}$$

$$\frac{\sum_{y_n h_t(\mathbf{x}_n)=-1} w_t(n)}{\sum_{n=1}^{N} w_t(n)} e^{\alpha_t} + \frac{\sum_{y_n h_t(\mathbf{x}_n)=1} w_t(n)}{\sum_{n=1}^{N} w_t(n)} e^{-\alpha_t}$$

$$\varepsilon_t \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}} + (1-\varepsilon_t) \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}} = 2\sqrt{\varepsilon_t(1-\varepsilon_t)} \leq 1$$

$$\sqrt{\varepsilon_t(1-\varepsilon_t)} \leq 1$$

$$1 - \varepsilon_t$$

$$\varepsilon_t = 1 - \varepsilon_t = 1/2$$

$$E_0 = \sum_{n=1}^{N} w_0(n) = N$$

$\varepsilon < 1/2$

$$\lim_{t\to\infty} E_t \approx \lim_{t\to\infty} \left(2\sqrt{\varepsilon(1-\varepsilon)}\right)^t E_0 = 0$$

$\mathbf{d}_i$

$$x_n = P_{\mathbf{d}_i}(\mathbf{x}_n) = \frac{\mathbf{x}_n^T \mathbf{d}_i}{||\mathbf{d}_i||}, \qquad (n = 1, \cdots, N)$$

T

$$h(x_n) = \begin{cases} +1 & \text{if } x_n < T \\ -1 & \text{if } x_n > T \end{cases}$$

$$\varepsilon_t = \sum_{n=1}^{N} w_t(n)\ \delta(h_t(x_n) - y_n) = \sum_{h_t(x_n) \neq y_n} w_t(n)$$

$$\mathbf{S}_b = \frac{1}{N}\left[N_-(\mathbf{m}_{-1}-\mathbf{m})(\mathbf{m}_{-1}-\mathbf{m})^T + N_+(\mathbf{m}_{+1}-\mathbf{m})(\mathbf{m}_{+1}-\mathbf{m})^T\right]$$

N_{-}

N_+

$$(N_- + N_+ = N)$$

$$\mathbf{m}_{\pm} = \frac{1}{N_{\pm}} \sum_{y_n = \pm 1} w(n)\, \mathbf{x}_n, \qquad \mathbf{m} = \frac{1}{N} \sum_{n=1}^{N} w(n)\, \mathbf{x}_n$$

S_b

$$\mathbf{S}_b = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^{-1} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^T$$

Φ

I

C_0

$$N = 200$$

$$N = 500$$

E_t

$$E_{t+1}/E_t = 2\sqrt{\varepsilon_t(1-\varepsilon_t)}$$

$\varepsilon_t < 0.5$

$$E_{t+1}/E_t < 1$$

$$\{(\mathbf{x}_n, y_n), \quad n = 1, \cdots, N\}$$

$$y_n = 1$$

$$y_n = -1$$

H_0

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = \sum_{i=1}^{d} x_i w_i + b = 0$$

W

b

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b \begin{cases} > 0 \\ = 0 \\ < 0 \end{cases},$$

$$\begin{cases} \mathbf{x} \in C_+ \\ \mathbf{x} \in H_0 \\ \mathbf{x} \in C_- \end{cases}$$

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$$

$$f(\mathbf{x}) = 0$$

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0$$

$$p_{\mathbf{w}}(\mathbf{x}) = \frac{\mathbf{w}^T\mathbf{x}}{||\mathbf{w}||} = -\frac{b}{||\mathbf{w}||}$$

$\mathbf{x} \in H_0$

$$d(H_0,\, \mathbf{0}) = |p_{\mathbf{w}}(\mathbf{x})| = \frac{|b|}{||\mathbf{w}||}$$

$$\mathbf{x}' \notin H_0$$

$$p_{\mathbf{w}}(\mathbf{x}') = \mathbf{w}^T\mathbf{x}'/||\mathbf{w}||$$

$\mathbf{x}'$

$$d(H_0,\ \mathbf{x}') = \left| p_{\mathbf{w}}(\mathbf{x}') - p_{\mathbf{w}}(\mathbf{x}) \right| = \left| \frac{\mathbf{w}^T\mathbf{x}'}{||\mathbf{w}||} - \left(-\frac{b}{||\mathbf{w}||} \right) \right| = \frac{|\mathbf{w}^T\mathbf{x}' + b|}{||\mathbf{w}||} = \frac{|f(\mathbf{x}')|}{||\mathbf{w}||}$$

$$\mathbf{x}_{sv}$$

$$d(H_0,\, \mathbf{x}_{sv} \in C_+) = d(H_0,\, \mathbf{x}_{sv} \in C_-) = d(H_0,\, \mathbf{x}_{sv}) = \frac{|f(\mathbf{x}_{sv})|}{||\mathbf{w}||}$$

$$H_+$$

H_{-}

$$\begin{cases} H_+ : & f_+(\mathbf{x}) = f(\mathbf{x}) - c = \mathbf{w}^T\mathbf{x} + b - c = 0 \\ H_- : & f_-(\mathbf{x}) = f(\mathbf{x}) + c = \mathbf{w}^T\mathbf{x} + b + c = 0 \end{cases}$$

$$c = 1$$

$$d(H_{\pm}, \mathbf{0}) = \frac{|b \pm 1|}{||\mathbf{w}||}$$

$$\left| d(H_{\pm}, \mathbf{0}) - d(H_0, \mathbf{0}) \right| = \left| \frac{|b \pm 1|}{||\mathbf{w}||} - \frac{|b|}{||\mathbf{w}||} \right| = \frac{1}{||\mathbf{w}||}$$

$$||\mathbf{w}||$$

$$\{(\mathbf{x}_n, y_n) \quad (n = 1, \cdots, N)\}$$

$$\begin{cases} f(\mathbf{x}_n) - 1 = \mathbf{w}^T\mathbf{x}_n + b - 1 \geq 0 & \text{if } y_n = 1 \\ f(\mathbf{x}_n) + 1 = \mathbf{w}^T\mathbf{x}_n + b + 1 \leq 0 & \text{if } y_n = -1 \end{cases}$$

$$y_n \, f(\mathbf{x}_n) = y_n(\mathbf{w}^T \mathbf{x}_n + b) \geq 1$$

$$\frac{1}{2}||\mathbf{w}||^2 = \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$y_n(\mathbf{x}_n^T \mathbf{w} + b) \geq 1, \quad (n = 1, \cdots, N)$$

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = [w_1, w_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b = [1, 2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 1 = x_1 + 2x_2 - 1 = 0$$

$$d(H_0, \mathbf{0}) = \frac{|b|}{||\mathbf{w}||} = \frac{1}{\sqrt{w_1^2 + w_2^2}} = \frac{1}{\sqrt{1^2 + 2^2}} = \frac{1}{\sqrt{5}} = 0.447$$

$$\mathbf{x}_0 = [0.5,\ 0.25]^T$$

$$f(\mathbf{x}_0) = 0.5 + 2 \times 0.25 - 1 = 0$$

$$\mathbf{x}_0$$

$$d(\mathbf{x}_0, H) = f(\mathbf{x}_0)/||\mathbf{w}|| = 0$$

$$\mathbf{x}_1 = [1,\ 0.25]^T$$

$$f(\mathbf{x}_1) = 1 + 2 \times 0.25 - 1 = 0.5 > 0$$

$\mathbf{x}_1$

$$d(\mathbf{x}_1, H) = f(\mathbf{x}_1)/||\mathbf{w}|| = |0.5|/\sqrt{5} = 0.2235$$

$$\mathbf{x}_2 = [0.5,\ 0]^T$$

$$f(\mathbf{x}_2) = 0.5 + 2 \times 0 - 1 = -0.5 < 0$$

$$\mathbf{x}_2$$

$$d(\mathbf{x}_2, H) = f(\mathbf{x}_2)/||\mathbf{w}|| = |-0.5|/\sqrt{5} = 0.2235$$

$$\mathbf{x}_1 = [1,\ 0.25]^T$$

$$\mathbf{x}_2 = [0.5,\ 0]^T$$

$$f(\mathbf{x}) = x_1 + 2x_2 - 1 = 0$$

$$|f(\mathbf{x}_{sv})| = |\mathbf{w}^T\mathbf{x}_1 + b| = |\mathbf{w}^T\mathbf{x}_2 + b| = 0.5$$

0.5

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 2x_1 + 4x_2 - 2 = 0$$

H_+ :

$$2x_1 + 4x_2 - 2 - 1 = 2x_1 + 4x_2 - 3 = 0$$

H_{-} :

$$2x_1 + 4x_2 - 2 + 1 = 2x_1 + 4x_2 - 1 = 0$$

$$d(H_+,\, \mathbf{0}) = \frac{|-3|}{||\mathbf{w}||} = 1.341, \quad d(H_-,\, \mathbf{0}) = \frac{|-1|}{||\mathbf{w}||} = 0.447$$

$$2/||\mathbf{w}||$$

$$d(H_-,\, H_+) = d(H_+,\, \mathbf{0}) - d(H_-,\, \mathbf{0}) = \frac{2}{||\mathbf{w}||} = 0.894$$

$$d(H_0, \mathbf{0}) = 0.447$$

$$L_p(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n(1 - y_n(\mathbf{w}^T\mathbf{x}_n + b))$$

$\alpha_1, \cdots, \alpha_N$

$$\alpha_n$$

$$\alpha_n \geq 0$$

$$y_n(\mathbf{w}^T\mathbf{x}_n + b) = 1$$

$$\alpha_n > 0$$

$$\alpha_n = 0$$

$$L_p(\mathbf{w}, b)$$

$$\frac{\partial}{\partial b}L_p(\mathbf{w},b) = \frac{\partial}{\partial b}\left[\frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N}\alpha_n(1-y_n(\mathbf{w}^T\mathbf{x}_n+b))\right] = \sum_{n=1}^{N}\alpha_n y_n = 0$$

$$\frac{\partial}{\partial \mathbf{w}} L_p(\mathbf{w}, b) = \frac{\partial}{\partial \mathbf{w}} \left[\frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{n=1}^{N} \alpha_n (1 - y_n(\mathbf{w}^T \mathbf{x}_n + b)) \right] = \mathbf{w} - \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n = 0,$$

$$\mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n$$

$$L_p(\mathbf{w}, b, \alpha)$$

$$\{\alpha_1, \cdots, \alpha_N\}$$

$$L_d(\alpha)$$

$$\inf_{\mathbf{w},b} L_p(\mathbf{w}, b, \alpha) = \inf_{\mathbf{w},b} \left[\frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n(1 - y_n(\mathbf{w}^T\mathbf{x}_n + b)) \right]$$

$$\inf_{\mathbf{w},b} \left[\frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n - \mathbf{w}^T \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n - b \sum_{n=1}^{N} \alpha_n y_n \right]$$

$$\frac{1}{2}\left(\sum_{m=1}^{N}\alpha_m y_m \mathbf{x}_m\right)^T\left(\sum_{n=1}^{N}\alpha_n y_n \mathbf{x}_n\right)+\sum_{n=1}^{N}\alpha_n-\left(\sum_{m=1}^{N}\alpha_m y_m \mathbf{x}_m\right)^T\left(\sum_{n=1}^{N}\alpha_n y_n \mathbf{x}_n\right)$$

$$\sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \left(\mathbf{x}_n^T \mathbf{x}_m \right)$$

$L_d(\alpha)$

$$L_d(\alpha) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \left(\mathbf{x}_n^T \mathbf{x}_m\right) = \mathbf{1}^T \alpha - \frac{1}{2} \alpha^T \mathbf{Q} \alpha$$

$$\sum_{n=1}^{N} \alpha_n y_n = \mathbf{y}^T \alpha = 0, \quad \alpha_n \geq 0 \quad (n = 1, \cdots, N)$$

Q

$$Q(m, n) = y_m y_n \mathbf{x}_m^T \mathbf{x}_n$$

$m, n = 1, \cdots, N$

$$\alpha_n \geq 0, \ (n = 1, \cdots, N)$$

$$\mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n = \sum_{n \in sv} \alpha_n y_n \mathbf{x}_n = \sum_{x_n \in C_+,\ n \in sv} \alpha_n \mathbf{x}_n - \sum_{x_n \in C_-,\ n \in sv} \alpha_n \mathbf{x}_n$$

$$y_n \left(\mathbf{w}^T \mathbf{x}_n + b \right) = 1$$

$$\mathbf{w}^T \mathbf{x}_n + b = y_n, \qquad (n \in sv)$$

$$y_n^2 = 1$$

$$b = y_n - \mathbf{w}^T \mathbf{x}_n = y_n - \sum_{m \in sv} \alpha_m y_m (\mathbf{x}_m^T \mathbf{x}_n), \qquad (n \in sv)$$

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = \left(\sum_{n\in sv} \alpha_n y_n \mathbf{x}_n\right)^T \mathbf{x} + b = \sum_{n\in sv} \alpha_n y_n \left(\mathbf{x}_n^T \mathbf{x}\right) + b \quad \begin{cases} > 0 & \mathbf{x} \in C_+ \\ < 0 & \mathbf{x} \in C_- \end{cases}$$

$$||\mathbf{w}||^2$$

$$\mathbf{w}^T\mathbf{w} = \left(\sum_{n\in sv} \alpha_n y_n \mathbf{x}_n^T\right) \left(\sum_{m\in sv} \alpha_m y_m \mathbf{x}_m\right) = \sum_{n\in sv} \alpha_n y_n \sum_{m\in sv} \alpha_m y_m (\mathbf{x}_n^T \mathbf{x}_m)$$

$$\sum_{n \in sv} \alpha_n y_n (y_n - b) = \sum_{n \in sv} \alpha_n (1 - y_n b) = \sum_{n \in sv} \alpha_n - b \sum_{n \in sv} \alpha_n y_n = \sum_{n \in sv} \alpha_n$$

$$\sum_{n=1}^{N} \alpha_n y_n = 0$$

$$\frac{1}{||\mathbf{w}||} = \left(\sum_{n \in sv} \alpha_n \right)^{-1/2}$$

y_n

$$\mathbf{x}_n^T \mathbf{x}_m$$

$$\left[\mathbf{x}_1, \cdots, \mathbf{x}_N\right]$$

$y_1, \cdots, y_N$

$$\mathbf{x}^T\mathbf{Q}\mathbf{x}/2 + \mathbf{c}^T\mathbf{x}$$

$$\mathbf{Ax} = \mathbf{b}$$

$$\mathbf{x} \geq \mathbf{0}$$

$$\mathbf{m}_- = \begin{bmatrix} 3.5 \\ 0 \end{bmatrix}, \quad \mathbf{S}_- = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \quad \mathbf{m}_+ = \begin{bmatrix} 0 \\ 3.5 \end{bmatrix}, \quad \mathbf{S}_+ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

n	α_n	$\mathbf{x}_n = [x_1,\ x_2]$	y_n
40	0.52	$x = [4.43,\ 2.44]$	-1
52	3.11	$x = [1.17,\ 0.74]$	-1
103	3.64	$x = [1.30,\ 1.64]$	1

,

$$\mathbf{w} = \begin{bmatrix} -1.25 \\ 2.39 \end{bmatrix}, \qquad b = -1.31$$

$$\mathbf{x} \Longrightarrow \mathbf{z} = \phi(\mathbf{x})$$

$$C_- = \{x \big| (\alpha \leq x \leq \beta)\}$$

$$C_+ = \{x \big| (x \leq \alpha)$$

$(x \geq \beta)\}$

$$\mathbf{z} = \phi(x) = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} x \\ (x - (\alpha + \beta)/2)^2 \end{bmatrix}$$

$$C_- = \{\mathbf{x},\ ||\mathbf{x}|| < d\}$$

$$C_+ = \{\mathbf{x},\ ||\mathbf{x}|| > d\}$$

$$\mathbf{z} = \phi(\mathbf{x}) = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_1^2 + x_2^2 \end{bmatrix}$$

$$\mathbf{z} = \phi(\mathbf{x}) = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_1 x_2 \end{bmatrix}$$

$\mathbf{x}_m$

$$\mathbf{z}_m = \phi(\mathbf{x}_m)$$

$$\mathbf{z}_n = \phi(\mathbf{x}_n)$$

$$K(\mathbf{x}_m, \mathbf{x}_n) = \phi(\mathbf{x}_m)^T \phi(\mathbf{x}_n) = \mathbf{z}_m^T \mathbf{z}_n$$

$$\mathbf{z}_m$$

$$\mathbf{z}_n$$

$$\mathbf{z} = \phi(\mathbf{x})$$

$$K(\mathbf{x}, \mathbf{x}')$$

$$\mathbf{z}' = \phi(\mathbf{x}')$$

$$\mathbf{x}' = [x'_1, \cdots, x'_d]^T$$

$$K(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}' = \sum_{i=1}^{d} x_i x_i'$$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!}\ x^{n-k} y^k$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

n

$$n - k$$

$$(x_1 + \cdots + x_d)^n = \sum_{\sum_{i=1}^d k_i = n} \begin{pmatrix} n \\ k_1, \cdots, k_d \end{pmatrix} x_1^{k_1} \cdots x_d^{k_d} = \sum_{\sum_{i=1}^d k_i = n} \frac{n!}{k_1! \cdots k_d!} x_1^{k_1} \cdots x_d^{k_d}$$

$$\begin{pmatrix} n \\ k_1, \cdots, k_d \end{pmatrix} = \frac{n!}{k_1! \cdots k_d!}$$

k_i

$k_1, \cdots, k_d$

$$K(\mathbf{x}, \mathbf{x}')$$

$$(\mathbf{x}^T\mathbf{x}')^n = (x_1x_1' + \cdots + x_dx_d')^n$$

$$\sum_{\sum_{i=1}^{d} k_i = n} \frac{n!}{k_1! \cdots k_d!} \left((x_1 x'_1)^{k_1} \cdots (x_d x'_d)^{k_d} \right) = \phi(\mathbf{x})^T \phi(\mathbf{x}') = \mathbf{z}^T \mathbf{z}'$$

$$\mathbf{z} = \phi(\mathbf{x}) = \left[\sqrt{\frac{n!}{k_1! \cdots k_d!}} \left(x_1^{k_1} \cdots x_d^{k_d} \right), \ \left(k_i \geq 0, \ \sum_{i=1}^{d} k_i = n \right) \right]^T$$

$$d = 2$$

$n = 2$

$$\mathbf{x} = [x_1, x_2]^T$$

$$K(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}')^2 = (x_1 x_1' + x_2 x_2')^2 = (x_1 x_1')^2 + 2 x_1 x_1' x_2 x_2' + (x_2 x_2')^2 = \phi(\mathbf{x})^T \phi(\mathbf{x}') = \mathbf{z}^T \mathbf{z}'$$

$$\mathbf{z} = \phi(\mathbf{x}) = [x_1^2,\ \sqrt{2}x_1x_2,\ x_2^2]$$

Z

$$K(\mathbf{x}, \mathbf{x}') = (1 + \mathbf{x}^T \mathbf{x}')^n$$

$$K(\mathbf{x}, \mathbf{x}') = e^{-||\mathbf{x}-\mathbf{x}'||^2/2\sigma^2} = e^{-\gamma||\mathbf{x}-\mathbf{x}'||^2}$$

$$\gamma = 1/2\sigma^2$$

$\sigma = 1$

$$K(\mathbf{x},\, \mathbf{x}')$$

$$e^{-||\mathbf{x}-\mathbf{x}'||^2/2} = e^{-||\mathbf{x}||^2/2}\ e^{-||\mathbf{x}'||^2/2}\ e^{\mathbf{x}^T\mathbf{x}'} = e^{-||\mathbf{x}||^2/2}\ e^{-||\mathbf{x}'||^2/2} \sum_{n=0}^{\infty} \frac{(\mathbf{x}^T\mathbf{x}')^n}{n!}$$

$$e^{-||\mathbf{x}||^2/2}\ e^{-||\mathbf{x}'||^2/2}\ \sum_{n=0}^{\infty}\left[\frac{1}{n!}\sum_{\sum_{i=1}^{d}k_i=n}\frac{n!}{k_1!\cdots k_d!}\left((x_1x_1')^{k_1}\cdots(x_dx_d')^{k_d}\right)\right]$$

$$\sum_{n=0}^{\infty} \sum_{\sum_{i=1}^{d} k_i = n} \left(e^{-||\mathbf{x}||^2/2} \; \frac{x_1^{k_1} \cdots x_d^{k_d}}{\sqrt{k_1! \cdots k_d!}} \right) \left(e^{-||\mathbf{x}'||^2/2} \; \frac{x_1'^{k_1} \cdots x_d'^{k_d}}{\sqrt{k_1! \cdots k_d!}} \right)$$

$$\phi(\mathbf{x})^T \phi(\mathbf{x}') = \mathbf{z}^T \mathbf{z}'$$

$$\mathbf{z} = \phi(\mathbf{x}) = \left[e^{-||\mathbf{x}||^2/2} \ \frac{x_1^{k_1} \cdots x_d^{k_d}}{\sqrt{k_1! \cdots k_d!}}, \ \left(n = 0, \cdots, \infty, \ \sum_{k=1}^{n} k_i = n \right) \right]^T$$

$$d = 1$$

$$K(x, x')$$

$$e^{-(x-x')^2/2} = e^{-x^2/2}\, e^{-x'^2/2}\, e^{xx'} = e^{-x^2/2}\, e^{-x'^2/2} \sum_{n=0}^{\infty} \frac{(xx')^n}{n!}$$

$$\sum_{n=0}^{\infty}(e^{-x^2/2}\,x^n/\sqrt{n!})\;(e^{-x'^2/2}\,x'^n/\sqrt{n!})$$

$$\mathbf{z} = \phi(x) = \left[e^{-x^2/2}\, x^n / \sqrt{n!},\ (n = 0, \cdots, \infty) \right]^T$$

$$\mathbf{x}_m^T \mathbf{x}_n$$

$$K(\mathbf{x}_m,\,\mathbf{x}_n)$$

$$L_d(\alpha) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m K(\mathbf{x}_n, \mathbf{x}_m)$$

$$b = y_n - \sum_{m \in sv} \alpha_m y_m K(\mathbf{x}_m, \mathbf{x}_n), \qquad (n \in sv)$$

$$\mathbf{x}_n^T \mathbf{x}$$

$$K(\mathbf{x}_n,\ \mathbf{x})$$

$$f(\mathbf{z}) = \mathbf{w}^T\mathbf{z} + b = \sum_{n\in sv} \alpha_n y_n \,(\mathbf{z}_n^T\mathbf{z}) + b = \sum_{n\in sv} \alpha_n y_n \, K(\mathbf{x}_n, \mathbf{x}) + b \qquad \begin{cases} > 0 & \mathbf{x} \in C_+ \\ < 0 & \mathbf{x} \in C_- \end{cases}$$

$$\zeta_n \geq 0$$

$$y_n(\mathbf{x}_n^T \mathbf{w} + b) \geq 1 - \xi_n, \quad (n = 1, \cdots, N)$$

ξ_n

$$\mathbf{w}^T\mathbf{w} + C\sum_{n=1}^{N}\xi_n$$

$$y_n(\mathbf{x}_n^T \mathbf{w} + b) + \xi_n - 1 \geq 0,$$

$$\xi_n \geq 0; \quad (n = 1, \cdots, N)$$

C

$$L_p(\mathbf{w}, b, \xi, \alpha, \mu) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}\alpha_n[y_n(\mathbf{w}^T\mathbf{x} + b) + \xi_n - 1] - \sum_{n=1}^{N}\mu_n\xi_n$$

$$\mu_n$$

$$\frac{\partial L_p}{\partial \mathbf{w}} = \mathbf{w} - \sum_{n=1}^{N} y_n \alpha_n \mathbf{x}_n = 0,$$

$$\mathbf{w} = \sum_{n=1}^{N} y_n \alpha_n \mathbf{x}_n;$$

$$\frac{\partial L_p}{\partial b} = \sum_{n=1}^{N} y_n \alpha_n = 0;$$

$$\frac{\partial L_p}{\partial \xi_n} = C - \alpha_n - \mu_n = 0;$$

$$y_n(\mathbf{w}^T\mathbf{x} + b) + \xi_n - 1 \geq 0, \quad \xi_n \geq 0,$$

$$\alpha_n[y_n(\mathbf{w}^T\mathbf{x} + b) + \xi_n - 1] = 0, \quad \mu_n\xi_n = 0,$$

$$\alpha_n \geq 0, \qquad \mu_n \geq 0$$

$$\mu_n = C - \alpha_n = C > 0$$

$$\xi_n = 0$$

$$y_n(\mathbf{w}^T \mathbf{x}_n + b) - 1 \geq 0$$

$$0 < \alpha_n < C$$

$\alpha_n \neq 0$

$$y_n(\mathbf{w}^T\mathbf{x}_n + b) + \xi_n - 1 = 0$$

$$\mu_n = C - \alpha_n > 0$$

$$y_n(\mathbf{w}^T\mathbf{x}_n + b) - 1 = 0$$

$$\alpha_n = C \neq 0$$

$$\mu_n = C - \alpha_n = 0$$

$$\xi_n \geq 0$$

$$y_n(\mathbf{w}^T\mathbf{x}_n + b) - 1 \leq 0$$

$$y_n(\mathbf{w}^T\mathbf{x}_n + b) - 1$$

$$y_n(\mathbf{w}^T\mathbf{x}_n + b) - 1 = y_n(\mathbf{w}^T\mathbf{x}_n + b) - y_n^2 = y_n[(\mathbf{w}^T\mathbf{x}_n + b) - y_n] = y_n\, E_n$$

$$E_n = (\mathbf{w}^T \mathbf{x}_n + b) - y_n$$

$$
y_n \, E_n \begin{cases} \geq 0 & \text{if } \alpha_n = 0 \\ = 0 & \text{if } 0 < \alpha_n < C \\ \leq 0 & \text{if } \alpha_n = C \end{cases}
$$

$$L_d(\alpha) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \mathbf{x}_m^T \mathbf{x}_n + C \sum_{n=1}^{N} \xi_n - \sum_{n=1}^{N} \alpha_n \xi_n - \sum_{n=1}^{N} \mu_n \xi_n$$

$$= \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \mathbf{x}_m^T \mathbf{x}_n$$

$$0 \leq \alpha_n \leq C, \quad \sum_{n=1}^{N} \alpha_n y_n = 0$$

$$C = \alpha_n + \mu_n$$

L_d

$$\alpha_n \geq 0, \ \mu_n \geq 0$$

$$0 \leq \alpha_n = C - \mu_n \leq C$$

$$\mathbf{w} = \sum_{n=1}^{N} y_n \alpha_n \mathbf{x}_n = \sum_{n \in sv} y_n \alpha_n \mathbf{x}_n$$

$$b = y_n - \sum_{i \in sv} y_i \alpha_i \mathbf{x}_i^T \mathbf{x}_n = y_n - \sum_{i \in sv} y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}_n), \qquad (n \in sv)$$

$$\mathbf{x}_k^T \mathbf{x}$$

$$K(\mathbf{x}_k, \mathbf{x})$$

$$\mathbf{w}^T\mathbf{x}+b=\sum_{i\in sv} y_i\alpha_i\mathbf{x}_i^T\mathbf{x}+b=\sum_{i\in sv} y_i\alpha_i K(\mathbf{x}_i,\mathbf{x})+b \left\{\begin{array}{ll} >0, & \mathbf{x}\in C_+ \\ <0, & \mathbf{x}\in C_- \end{array}\right.$$

$\alpha_1, \cdots, \alpha_{i-1}, \alpha_{i+1}, \cdots, \alpha_N$

$N-1$

α_j

$$N-2$$

$$L(\alpha_i, \alpha_j) = \alpha_i + \alpha_j - \frac{1}{2}\left(\alpha_i^2 \mathbf{x}_i^T \mathbf{x}_i + \alpha_j^2 \mathbf{x}_j^T \mathbf{x}_j + 2\alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j\right)$$

$$-\alpha_i y_i \left(\sum_{n \neq i} \alpha_n y_n \mathbf{x}_n^T \right) \mathbf{x}_i - \alpha_j y_j \left(\sum_{n \neq j} \alpha_n y_n \mathbf{x}_n^T \right) \mathbf{x}_j$$

$$\alpha_i + \alpha_j - \frac{1}{2}\left(\alpha_1^2 K_{ii} + \alpha_2^2 K_{jj} + 2\alpha_i \alpha_j y_i y_j K_{ij}\right)$$

$$-\alpha_i y_i \sum_{n \neq i} \alpha_n y_n K_{ni} - \alpha_j y_j \sum_{n \neq j} \alpha_n y_n K_{nj}$$

$$0 \leq \alpha_i, \alpha_j \leq C, \qquad \sum_{n=1}^{N} \alpha_n y_n = 0$$

$$K_{mn} = K(\mathbf{x}_m, \mathbf{x}_n) = \mathbf{z}_m^T \mathbf{z}_n$$

$\alpha_n^{old}\,(n = 1, \cdots, N)$

$$\alpha_i^{new}$$

$$\alpha_j^{new}$$

$$L(\alpha_i, \alpha_j)$$

$$\alpha_i y_i + \alpha_j y_j = -\sum_{n \neq i,j} \alpha_n y_n$$

y_i

$$y_i^2\alpha_i + y_i y_j \alpha_j = \alpha_i + s\alpha_j = \left(-\sum_{n\neq i,j} \alpha_n y_n \right) y_i = \delta,$$

$$\alpha_i = \delta - s\alpha_j$$

$$s = y_i y_j$$

$$\delta = (-\sum_{n \neq i,j} \alpha_n y_n) y_i$$

$$\alpha_i^{new} + s\alpha_j^{new} = \alpha_i^{old} + s\alpha_j^{old} = \delta,$$

$$\Delta\alpha_i = \alpha_i^{new} - \alpha_i^{old} = -s(\alpha_j^{new} - \alpha_j^{old}) = -s\Delta\alpha_j$$

$$\sum_{n \neq i,j} \alpha_n y_n \mathbf{x}_n = \mathbf{w} - \alpha_i y_i \mathbf{x}_i - \alpha_j y_j \mathbf{x}_j$$

$v_k \ (k = 1, 2)$

$$v_k$$

$$\sum_{n\neq i,j} \alpha_n y_n K_{nk} = \sum_{n\neq i,j} \alpha_n y_n K_{nk} - \alpha_i y_i K_{ik} - \alpha_j y_j K_{jk}$$

$$\left(\sum_{n\neq i,j}\alpha_n y_n K_{nk} + b\right) - b - \alpha_i y_i K_{ik} - \alpha_j y_j K_{jk}$$

$$u_k - b - \alpha_i y_i K_{ik} - \alpha_j y_j K_{jk}$$

$$u_k = \sum_{n=1}^{N} \alpha_n y_n K_{nk} + b$$

$$\mathbf{x}_k$$

$$L(\alpha_i, \alpha_j)$$

$$\alpha_i + \alpha_j - \frac{1}{2}(\alpha_i^2 K_{ii} + \alpha_j^2 K_{jj} + 2s\alpha_i\alpha_j K_{ij}) - \alpha_i y_i v_i - \alpha_j y_j v_j$$

$$\alpha_i = \delta - s\alpha_j$$

$$L(\alpha_j)$$

$$\delta + (1-s)\alpha_j - \frac{1}{2}(\delta - s\alpha_j)^2 K_{ii} - \frac{1}{2}\alpha_2^2 K_{jj}$$

$$-s(\delta - s\alpha_2)\alpha_2 K_{ij} - (\delta - s\alpha_j)y_i v_i - \alpha_j y_j v_j$$

$$\frac{1}{2}(2K_{ij} - K_{ii} - K_{jj})\alpha_j^2 + [1 - s + s\delta(K_{ii} - K_{ij}) + y_j(v_i - v_j)]\alpha_j + c$$

c

$$\alpha_j^2$$

$$\frac{1}{2}(2K_{ij} - K_{ii} - K_{jj}) = \frac{1}{2}\eta$$

η

η

$$2K_{ij} - K_{ii} - K_{jj} = 2\mathbf{z}_i^T \mathbf{z}_j - \mathbf{z}_i^T \mathbf{z}_i - \mathbf{z}_j^T \mathbf{z}_j = -(\mathbf{z}_i - \mathbf{z}_j)^T (\mathbf{z}_i - \mathbf{z}_j)$$

$$-||\mathbf{z}_i - \mathbf{z}_j||^2 \leq 0$$

$$1 - s + s\delta(K_{ii} - K_{ij}) + y_j(v_i - v_j)$$

$$1 - s + (s\alpha_i + \alpha_j)(K_{ii} - K_{ij})$$

$$+y_j[(u_i - b - \alpha_i y_i K_{ii} - \alpha_j y_j K_{ij}) - (u_j - b - \alpha_i y_i K_{ij} - \alpha_j y_j K_{jj})]$$

$$1 - s + s\alpha_1(K_{ii} - K_{ij}) + \alpha_2(K_{ii} - K_{ij})$$

$$+y_j(u_i - u_j) - \alpha_i s K_{ii} - \alpha_j K_{ij} + \alpha_i s K_{ij} + \alpha_j K_{jj}$$

$$y_j^2 - s - \alpha_j(2K_{ij} - K_{ii} - 2K_{jj}) + y_j(u_i - u_j)$$

$$y_j(u_i - y_j - u_j + y_j) - \alpha_j(2K_{ij} - K_{ii} - K_{jj})$$

$$y_j(E_i - E_j) - \alpha_j \eta$$

$$E_k = u_k - y_k = \sum_{n=1}^{N} \alpha_n^{old} y_n K_{nk} + b - y_k \qquad (k = 1, 2)$$

y_k

$\alpha_n^{old}\ (n = 1, \cdots, N)$

$$L(\alpha_j) = \frac{1}{2}\eta\alpha_j^2 + \left[y_j(E_i - E_j) - \alpha_j^{old}\eta\right]\alpha_j$$

$$L(\alpha_j)$$

$$\frac{d}{d\alpha_j} L(\alpha_j)$$

$$\eta\alpha_j + y_j(E_i - E_j) - \alpha_j^{old}\eta$$

$$\frac{d^2}{d\alpha_j^2}L(\alpha_j)$$

$$\eta \leq 0$$

$\eta \leq 0$

$$d\,L(\alpha_j)/d\alpha_j = 0$$

α_j^{old}

$$\alpha_j^{new} = \alpha_j^{old} + \frac{y_j(E_j - E_i)}{\eta} = \alpha_j^{old} + \Delta\alpha_j, \qquad \Delta\alpha_j = \alpha_j^{new} - \alpha_j^{old} = \frac{y_j(E_j - E_i)}{\eta},$$

$$0 \leq \alpha_i,\ \alpha_j \leq C$$

$$\alpha_i + s\alpha_j = \delta$$

$$s = 1$$

$$y_i = y_j = \pm 1$$

$$s = -1$$

$$y_i = -y_j = \pm 1$$

$$\delta < C$$

$$\delta > C$$

L

H

α_2

$$\alpha_i + s\alpha_j = \alpha_i + \alpha_j = \delta$$

$$\min(\alpha_j) = 0,\ \max(\alpha_j) = \delta$$

$$\min(\alpha_j) = \delta - C,\ \max(\alpha_j) = C$$

$$\begin{cases} L = \max(0,\ \delta - C) = \max(0,\ \alpha_i + \alpha_j - C) \\ H = \min(\delta,\ C) = \min(\alpha_i + \alpha_j,\ C) \end{cases}$$

$$\alpha_i + s\alpha_j = \alpha_i - \alpha_j = \delta$$

$\delta > 0$

$$\min(\alpha_j) = 0,\ \max(\alpha_j) = C - \delta$$

$\delta < 0$

$$\min(\alpha_j) = -\delta, \ \max(\alpha_j) = C$$

$$\begin{cases} L = \max(0,\ -\delta) = \max(0,\ \alpha_j - \alpha_i) \\ H = \min(C - \delta,\ C) = \min(C + \alpha_j - \alpha_i,\ C) \end{cases}$$

$$\alpha_j^{new} \Longleftarrow \begin{cases} H & \text{if } \alpha_j^{new} \geq H \\ \alpha_j^{new} & \text{if } L < \alpha_2^{new} < H \\ L & \text{if } \alpha_j^{new} \leq L \end{cases}$$

$$\alpha_i^{new} = \alpha_i^{old} - s\Delta\alpha_j = \alpha_i^{old} - s(\alpha_j^{new} - \alpha_j^{old})$$

$$\mathbf{w}^{new}$$

$$\sum_{n \neq i,j} y_n \alpha_n^{old} \mathbf{x}_n + y_i \alpha_i^{new} \mathbf{x}_i + y_j \alpha_j^{new} \mathbf{x}_j$$

$$\mathbf{w}^{old} - y_i\alpha_i^{old}\mathbf{x}_i - y_j\alpha_j^{old}\mathbf{x}_j + y_i\alpha_i^{new}\mathbf{x}_i + y_j\alpha_j^{new}\mathbf{x}_j$$

$$\mathbf{w}^{old} + y_i \Delta\alpha_i \mathbf{x}_i + y_j \Delta\alpha_j \mathbf{x}_j = \mathbf{w}^{old} + \Delta\mathbf{w}$$

$$\Delta \mathbf{w} = \mathbf{w}^{new} - \mathbf{w}^{old} = y_i \Delta\alpha_i \mathbf{x}_i + y_j \Delta\alpha_j \mathbf{x}_j$$

$$\Delta E_k$$

$$E_k^{new} - E_k^{old} = u_k^{new} - u_k^{old}$$

$$\sum_{n=1}^{N} \alpha_n^{new} y_n K_{nk} + b^{new} - \sum_{n=1}^{N} \alpha_n^{old} y_n K_{nk} - b^{old}$$

$$y_i \Delta\alpha_i K_{ik} + y_j \Delta\alpha_j K_{jk} + b^{new} - b^{old} \qquad (k = 1, 2)$$

$0 < \alpha_k < C$

$$k = i$$

$$k = j$$

$$y_k E_k = 0$$

$$E_k^{new} = 0$$

$$b_k^{new} = b_k^{old} - (E_k^{old} + y_i \Delta\alpha_i K_{ik} + y_j \Delta\alpha_j K_{jk}) \qquad (k = 1$$

2)

$$b_i^{new} = b_j^{new}$$

$$\alpha_k = 0$$

$$\alpha_k = C$$

$$k = 2$$

$$E_k \neq 0$$

$$b_i^{new} \neq b_j^{new}$$

$$b^{new} = (b_i^{new} + b_j^{new})/2$$

$$\Delta\alpha_n = \alpha_n^{new} - \alpha_n^{old}$$

$$yE = y_n E_n$$

tol

$m \times n$

$$\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$$

$$K(\mathbf{x}_i, \mathbf{x}),\ (i = 1, \cdots, n)$$

$C = 1.2$

$$C = 0.05$$

$$\sum_{n=1}^{N} \xi_n$$

$C = 0.1$

$K > 2$

$$C_+ = C_i$$

$$C_- = \{C_j | j = 1, \cdots, K,\ j \neq i\}$$

$$f(\mathbf{x})$$

$$f_i(\mathbf{x}) \begin{cases} > 0 \\ < 0 \end{cases},$$

$$\begin{cases} \mathbf{x} \in C_+ \\ \mathbf{x} \in C_- \end{cases}$$

$K - 1$

$i = 1, \cdots, K$

$$f_k(\mathbf{x}) = \max\{f_1(\mathbf{x}), \cdots, f_K(\mathbf{x})\}, \qquad (k = 1, \cdots, K)$$

$K(K-1)/2$

$$f_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + b_i \qquad (i = 1, \cdots, K)$$

$$\mathbf{x} = [x_0 = 1,\ x_1, \cdots, x_n]^T$$

$$\mathbf{w}_i = [w_{i0} = b_i,\ w_{i1}, \cdots, w_{in}]^T$$

$$f_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x}$$

$$\mathbf{f}(\mathbf{x}) = \mathbf{W}^T \mathbf{x}$$

$$\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_K]$$

$$f_i(\mathbf{x}) = \max(f_1(\mathbf{x}), \cdots, f_K(\mathbf{x}))$$

W

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{W} \mathbf{x} + \mathbf{w}^T \mathbf{x} + w \begin{cases} > 0, & \mathbf{x} \in C_+ \\ < 0, & \mathbf{x} \in C_- \end{cases}$$

$$-\frac{1}{2}(\mathbf{\Sigma}_{+}^{-1} - \mathbf{\Sigma}_{-}^{-1})$$

$$\boldsymbol{\Sigma}_{+}^{-1}\mathbf{m}_{+} - \boldsymbol{\Sigma}_{-}^{-1}\mathbf{m}_{-}$$

$$-\frac{1}{2}(\mathbf{m}_+^T \mathbf{\Sigma}_+^{-1} \mathbf{m}_+ - \mathbf{m}_-^T \mathbf{\Sigma}_-^{-1} \mathbf{m}_-) - \frac{1}{2} \log \frac{|\mathbf{\Sigma}_+|}{|\mathbf{\Sigma}_-|} + \log \frac{P(C_+)}{P(C_-)}$$

$$\Sigma_+$$

Σ_

$$\mathbf{m}_+ = \frac{1}{N_+} \sum_{x \in C_+} \mathbf{x}, \qquad \mathbf{m}_- = \frac{1}{N_-} \sum_{x \in C_-} \mathbf{x}$$

$$P_+ = N_+/N$$

$$P_{-} = N_{-}/N$$

$$\boldsymbol{\Sigma}_+ = \boldsymbol{\Sigma}_- = \boldsymbol{\Sigma}$$

$$\mathbf{W} = \mathbf{0}$$

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w = \left[\boldsymbol{\Sigma}^{-1} (\mathbf{m}_+ - \mathbf{m}_-) \right]^T \mathbf{x} + w = 0$$

$-w$

$$\boldsymbol{\Sigma}_{+}^{-1}$$

$$\boldsymbol{\Sigma}_{-}^{-1}$$

$$\mathbf{W} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{U}\mathbf{U}^T, \quad \mathbf{\Sigma}_+^{-1} = \mathbf{V}_+\mathbf{\Lambda}_+\mathbf{V}_+^T = \mathbf{U}_+\mathbf{U}_+^T, \quad \mathbf{\Sigma}_-^{-1} = \mathbf{V}_-\mathbf{\Lambda}_-\mathbf{V}_-^T = \mathbf{U}_-\mathbf{U}_-^T$$

$$\mathbf{U} = \mathbf{V}\boldsymbol{\Lambda}^{1/2},\ \mathbf{U}_{+} = \mathbf{V}_{+}\boldsymbol{\Lambda}_{+}^{1/2}$$

$$\mathbf{U}_{-} = \mathbf{V}_{-}\mathbf{\Lambda}_{-}^{1/2}$$

$$\mathbf{w} = \boldsymbol{\Sigma}_+^{-1}\mathbf{m}_+ - \boldsymbol{\Sigma}_-^{-1}\mathbf{m}_- = \mathbf{U}_+\mathbf{U}_+^T \frac{1}{N_+} \sum_{\mathbf{x}_+ \in C_+} \mathbf{x}_+ - \mathbf{U}_-\mathbf{U}_-^T \frac{1}{N_-} \sum_{\mathbf{x}_- \in C_-} \mathbf{x}_-$$

$$f(\mathbf{x})$$

$$\mathbf{x}^T \mathbf{W} \mathbf{x} + \mathbf{w}^T \mathbf{x} + w$$

$$\mathbf{x}^T\mathbf{U}\mathbf{U}^T\mathbf{x} + \left(\frac{1}{N_+}\sum_{\mathbf{x}_+}\mathbf{x}_+^T\mathbf{U}_+\mathbf{U}_+^T\right)\mathbf{x} - \left(\frac{1}{N_-}\sum_{\mathbf{x}_-}\mathbf{x}_-^T\mathbf{U}_-\mathbf{U}_-^T\right)\mathbf{x} + w$$

$$\mathbf{z}^T\mathbf{z} + \frac{1}{N_+}\sum_{\mathbf{z}_{++}}\left(\mathbf{z}_{++}^T\mathbf{z}_+\right) - \frac{1}{N_-}\sum_{\mathbf{z}_{--}}\left(\mathbf{z}_{--}^T\mathbf{z}_-\right) + w$$

$$\left\{ \begin{array}{l} \mathbf{z}_{++} = \mathbf{U}_+^T \mathbf{x}_+ \quad (\mathbf{x}_+ \in C_+) \\ \mathbf{z}_{--} = \mathbf{U}_-^T \mathbf{x}_- \quad (\mathbf{x}_- \in C_-) \end{array} \right. , \qquad \left\{ \begin{array}{l} \mathbf{z} = \mathbf{U}^T \mathbf{x} \\ \mathbf{z}_+ = \mathbf{U}_+^T \mathbf{x} \\ \mathbf{z}_- = \mathbf{U}_-^T \mathbf{x} \end{array} \right.$$

$$f(\mathbf{x}) = K(\mathbf{z},\mathbf{z}) + \frac{1}{N_+}\sum_{\mathbf{z}_{++}} K(\mathbf{z}_{++},\mathbf{z}_+) - \frac{1}{N_-}\sum_{\mathbf{z}_{--}} K(\mathbf{z}_{--},\mathbf{z}_-) + b = p(\mathbf{x}) + b$$

w

$$x_n = p(\mathbf{x}_n),\ (n = 1, \cdots, N)$$

$1, \cdots, k$

$k+1, \cdots, N$

$$\left|\sum_{n=1}^{k} y_n\right| + \left|\sum_{n=k+1}^{N} y_n\right|, \qquad (k = 1, \cdots, N-1)$$

$$(x_k + x_{k+1})/2$$

x_k

x_{k+1}

$$b = -(x_k + x_{k+1})/2$$

$$p(\mathbf{x}) + b \begin{cases} > 0, & \mathbf{x} \in C_+ \\ < 0, & \mathbf{x} \in C_- \end{cases}$$

$$\mathbf{W} = -(\mathbf{\Sigma}_{+} - \mathbf{\Sigma}_{-})/2 = \mathbf{0}$$

$$\mathbf{\Sigma} = (\mathbf{\Sigma}_+ + \mathbf{\Sigma}_-)/2$$

$$\mathbf{w}^T\mathbf{x} + w = \left[\boldsymbol{\Sigma}^{-1}(\mathbf{m}_+ - \mathbf{m}_-)\right]^T \mathbf{x} + w$$

$$\left[\mathbf{U}\mathbf{U}^T\left(\frac{1}{N_+}\sum_{\mathbf{x}_+}\mathbf{x}_+ - \frac{1}{N_-}\sum_{\mathbf{x}_-}\mathbf{x}_-\right)\right]^T \mathbf{x} + w$$

$$\frac{1}{N_+}\sum_{\mathbf{x}_+}\mathbf{x}_+^T\mathbf{U}\mathbf{U}^T\mathbf{x}-\frac{1}{N_-}\sum_{\mathbf{x}_-}\mathbf{x}_-^T\mathbf{U}\mathbf{U}^T\mathbf{x}+w$$

$$\frac{1}{N_+}\sum_{\mathbf{z}_+}\mathbf{z}_+^T\mathbf{z}-\frac{1}{N_-}\sum_{\mathbf{z}_-}\mathbf{z}_-^T\mathbf{z}+w$$

$$\mathbf{U}\mathbf{U}^T = \mathbf{\Sigma}^{-1}$$

$$\mathbf{z}_+ = \mathbf{U}^T \mathbf{x}_+$$

$$\mathbf{z}_- = \mathbf{U}^T \mathbf{x}_-$$

$$\mathbf{z} = \mathbf{U}^T \mathbf{x}$$

$$f(\mathbf{x}) = \frac{1}{N_+} \sum_{\mathbf{z}_+} K(\mathbf{z}_+, \mathbf{z}) - \frac{1}{N_-} \sum_{\mathbf{z}_-} K(\mathbf{z}_-, \mathbf{z}) + b$$

$$\boldsymbol{\Sigma}_+ = \boldsymbol{\Sigma}_- = \mathbf{I}$$

$$\mathbf{w}^T \mathbf{x} + w = (\mathbf{m}_+ - \mathbf{m}_-)^T \mathbf{x} + w$$

$$\frac{1}{N_+}\sum_{\mathbf{x}_+}\mathbf{x}_+^T\mathbf{x} - \frac{1}{N_-}\sum_{\mathbf{x}_-}\mathbf{x}_-^T\mathbf{x} + w$$

$$f(\mathbf{x}) = \frac{1}{N_+} \sum_{\mathbf{x}_+} K(\mathbf{x}_+, \mathbf{x}) - \frac{1}{N_-} \sum_{\mathbf{x}_-} K(\mathbf{x}_-, \mathbf{x}) + b$$

$$K(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}')$$

$$\phi(\mathbf{x}_n), \quad (n = 1, \cdots, N)$$

$$\phi(\mathbf{m}_+)$$

$$\phi(\mathbf{m}_-)$$

$$\phi(\mathbf{m}_+ - \mathbf{m}_-) \neq \phi(\mathbf{m}_+) - \phi(\mathbf{m}_-), \qquad \phi(\mathbf{m}_\pm) = \phi\left(\frac{1}{n_\pm}\sum_{\mathbf{x}\in C_\pm}\mathbf{x}\right) \neq \frac{1}{n_\pm}\sum_{\mathbf{x}\in C_\pm}\phi(\mathbf{x})$$

$$f(\mathbf{x}) = \mathbf{z}^T\mathbf{z} + \frac{1}{N_+}\sum_{\mathbf{z}_{++}} K(\mathbf{z}_{++}, \mathbf{z}_+) - \frac{1}{N_-}\sum_{\mathbf{z}_{--}} K(\mathbf{z}_{--}, \mathbf{z}_-) + b$$

$$\mathbf{y} = [y_1, \cdots, y_N]$$

$$K(\mathbf{x}_i, \mathbf{x}), \ (i = 1, \cdots, N)$$

X

$$\mathbf{W} = -(\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_1^{-1})/2$$

D

$$\mathbf{D}^{1/2}$$

$y > 0$

$y < 0$

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

γ

$$1 < \gamma < 9$$

	Kernel	Matlab SVM	Kernel Bayes I	Kernel Bayes II	Kernel Bayes III
Test 1	linear	93.0%	88.0%	94.0%	94.0%
Test 2	linear	73.0%	80.0%	79.5%	96.5%
Test 3	RBF	98.5%	100%	100%	100%

$$\mathbf{m}_{+} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \qquad \mathbf{m}_{-} = \begin{bmatrix} +1 \\ 0 \end{bmatrix}, \qquad \mathbf{\Sigma}_{+} = \mathbf{\Sigma}_{-} = 3 \times \begin{bmatrix} 1 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

(94%)

(93%)

(84.7%)

(96.5%)

(98.5%)

$$\gamma = 5$$

	Kernel	Matlab SVM	Kernel Bayes I	Kernel Bayes II	Kernel Bayes III
Test 1	Linear	58.75%	60.0%	60.25%	97.0%
Test 2	RBF	97.75%	98.50%	98.50%	99.50%
Test 3	RBF	98.0%	100%	100%	100%
Test 4	RBF	96.0%	98.5%	95.5%	97.0%

$$y = f(\mathbf{x}) = \mathbf{x}^T \mathbf{w}$$

$$\mathcal{D} = \{(\mathbf{x}_n,\, y_n) | n = 1, \cdots, N\}$$

$$y_n = 0$$

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{w}$$

$$\mathcal{D} = \{(\mathbf{x}_n, y_n) | n = 1, \cdots, N\}$$

$\mathcal{D}$

$$f(\mathbf{x}_*)$$

$\mathbf{x}_*$

$$p(y_* = 1|\mathbf{x}_*, \mathcal{D}) = \sigma(f(\mathbf{x}_*)) = \sigma(\mathbf{x}_*^T \mathbf{w})$$

$$p(y_* = 1 | \mathbf{x}_*, \mathcal{D})$$

$$\sigma(f(\mathbf{x}_*))$$

$$p(y_* = 1|\mathbf{x}_*, \mathcal{D})$$

$$\int p(y_* = 1|\mathbf{x}_*,\, f)\; p(f|\mathbf{x}_*, \mathcal{D})\, df$$

$$\int \sigma(f(\mathbf{x}_*))\, p(f|\mathbf{x}_*, \mathcal{D})\, df = E_f[\sigma(f(\mathbf{x}_*))] = \sigma(E_f f(\mathbf{x}_*))$$

$$\mathbf{X}_* = [\mathbf{x}_{1*}, \cdots, \mathbf{x}_{M*}]$$

$$p(\mathbf{y}_* = 1|\mathbf{X}_*, \mathcal{D})$$

$$\int p(\mathbf{y}_* = \mathbf{1} | \mathbf{X}_*, \, \mathbf{f}) \, p(\mathbf{f} | \mathbf{X}_*, \mathcal{D}) \, d\mathbf{f}$$

$$\int \sigma(\mathbf{f}(\mathbf{X}_*))\, p(\mathbf{f}|\mathbf{X}_*, \mathcal{D})\, d\mathbf{f} = E_f[\sigma(\mathbf{f}(\mathbf{X}_*))] = \sigma(E_f f(\mathbf{X}_*))$$

$$\mathbf{f}(\mathbf{X}_*) = [f(\mathbf{x}_{1*}), \cdots, f(\mathbf{x}_M)]^{T*}$$

$$p(\mathbf{f}|\mathbf{X}_*, \mathcal{D})$$

$$\mathbf{X}_*$$

$$p(\mathbf{f}|\mathcal{D})$$

$$\mathbf{f} = \mathbf{f}(\mathbf{X})$$

f

$$p(\mathbf{f}|\mathcal{D}) = p(\mathbf{f}|\mathbf{X}, \mathbf{y}) = \frac{p(\mathbf{y}, \mathbf{f}|\mathbf{X})}{p(\mathbf{y}|\mathbf{X})} = \frac{p(\mathbf{y}|\mathbf{f})\, p(\mathbf{f}|\mathbf{X})}{p(\mathbf{y}|\mathbf{X})} \propto p(\mathbf{y}|\mathbf{f})\, p(\mathbf{f}|\mathbf{X})$$

$$\mathbf{f}(\mathbf{X}) = [f(\mathbf{x}_1), \cdots, f(\mathbf{x}_N)]^T$$

$$p(\mathbf{y}|\mathbf{f}) = L(\mathbf{f}|\mathbf{y})$$

$$p(\mathbf{f}|\mathbf{X})$$

$$p(\mathbf{y}|\mathbf{X})$$

$$p(\mathbf{y}|\mathbf{f})$$

$$p(y = 1|\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x})$$

$$y = 1$$

$$y = -1$$

$$\mathbf{x} \in C_+$$

$\mathbf{x} \in C_{-}$

$$p(y = 1|f(\mathbf{x}))$$

$$\sigma(f(\mathbf{x}))$$

$$p(y = -1|f(\mathbf{x}))$$

$$1 - p(y = 1|f(\mathbf{x})) = 1 - \sigma(f(\mathbf{x})) = \sigma(-f(\mathbf{x}))$$

y

$$p(y|f(\mathbf{x})) = \sigma(y\, f(\mathbf{x})) = \sigma(y\, f)$$

$$p(\mathbf{y}|\mathbf{f}) = \prod_{n=1}^{N} p(y_n|f_n) = \prod_{n=1}^{N} \sigma(y_n\, f_n)$$

$$\mathbf{f}(\mathbf{X})$$

$$p(\mathbf{f}|\mathbf{X}) = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_f)$$

$$\Sigma_f$$

$$cov(f_m, f_n) = k(\mathbf{x}_m, \mathbf{x}_n) = \exp\left(-\frac{1}{a^2}||\mathbf{x}_m - \mathbf{x}_n||^2\right), \qquad (m, n = 1, \cdots, N)$$

$$f_m = f(\mathbf{x}_m)$$

$$f_n = f(\mathbf{x}_n)$$

a

$a \to \infty$

$$f(\mathbf{x}_m)$$

$$f(\mathbf{x}_n)$$

$$a \to 0$$

$$p(\mathbf{f}|\mathcal{D})$$

$\propto$

$$p(\mathbf{y}|\mathbf{f})\, p(\mathbf{f}|\mathbf{X}) = p(\mathbf{y}|\mathbf{f})\, \mathcal{N}(\mathbf{0}, \mathbf{K}) = \prod_{n=1}^{N} \sigma(y_n\, f_n)\ \frac{1}{(2\pi)^{d/2}|\mathbf{K}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{f}^T\mathbf{K}^{-1}\mathbf{f}\right)$$

$$\prod_{n=1}^{N} \sigma(y_n \, f_n) \ \exp\left(-\frac{1}{2}\mathbf{f}^T\mathbf{K}^{-1}\mathbf{f}\right)$$

$\mathbf{y}$

$$p(\mathbf{f}|\mathcal{D}) \approx \mathcal{N}(\mathbf{m}_{f|D}, \mathbf{\Sigma}_{f|D})$$

$$\mathbf{m}_{f|D}$$

$\Sigma_{f|D}$

$$\psi(\mathbf{f})$$

$$\psi(\mathbf{f}) = \log\, p(\mathbf{f}|\mathcal{D}) = \sum_{n=1}^{N} \log \sigma(y_n\, f_n) - \frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{K}| - \frac{1}{2}\mathbf{f}^T\mathbf{K}^{-1}\mathbf{f}$$

$$\psi(\mathbf{f}) = \log p(\mathbf{y}|\mathbf{f})$$

$$\mathbf{g}_{\psi}(\mathbf{f}) = \frac{d}{d\mathbf{f}}\psi(\mathbf{f}) = \frac{d}{d\mathbf{f}}\ \log\, p(\mathbf{y}|\mathbf{f})\ -\ \frac{d}{d\mathbf{f}}\left(\frac{1}{2}\mathbf{f}^T\mathbf{K}^{-1}\mathbf{f}\right) = \mathbf{w} - \mathbf{K}^{-1}\mathbf{f}$$

$$\mathbf{H}_\psi(\mathbf{f}) = \frac{d^2}{d\mathbf{f}^2}\psi(\mathbf{f}) = \frac{d}{d\mathbf{f}}\mathbf{g}_\psi(\mathbf{f}) = \frac{d}{d\mathbf{f}}\left(\mathbf{w} - \mathbf{K}^{-1}\mathbf{f}\right) = \mathbf{W} - \mathbf{K}^{-1}$$

$$\mathbf{w} = \frac{d}{d\mathbf{f}} \log p(\mathbf{y}|\mathbf{f}), \qquad \mathbf{W} = \frac{d^2}{d\mathbf{f}^2} \log p(\mathbf{y}|\mathbf{f}) = \frac{d\mathbf{w}}{d\mathbf{f}}$$

$$\frac{d}{df_n} \log\, p(y_n|f_n)$$

$$\frac{d}{df_n} \log\left(1 + \exp(-y_n f_n)\right)^{-1} = \frac{y_n}{1 + \exp(y_n f_n)}$$

$$\frac{d^2}{df_n^2} \log\, p(y_n|f_n)$$

$$\frac{d}{df_n}\left[\frac{y_n}{1+\exp(y_n f_n)}\right] = \frac{-\exp(-y_n f_n)}{(1+\exp(-y_n f_n))^2}$$

$$y_n = \pm 1$$

$$\mathbf{w} = \frac{d}{d\mathbf{f}} \log p(\mathbf{y}|\mathbf{f}) = \begin{bmatrix} d/df_1 \\ \vdots \\ d/df_N \end{bmatrix} \sum_{n=1}^{N} \log p(y_n|f_n) = \begin{bmatrix} d \log p(y_1|f_1)/df_1 \\ \vdots \\ d \log p(y_N|f_N)/df_N \end{bmatrix} = \begin{bmatrix} \frac{y_1}{1+e^{y_1 f_1}} \\ \vdots \\ \frac{y_N}{1+e^{y_N f_N}} \end{bmatrix}$$

$$\frac{d^2}{d\mathbf{f}^2}\log p(\mathbf{y}|\mathbf{f}) = \begin{bmatrix} \frac{\partial^2}{\partial f_1 \partial f_1} & \cdots & \frac{\partial^2}{\partial f_1 \partial f_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial f_N \partial f_1} & \cdots & \frac{\partial^2}{\partial f_N \partial f_N} \end{bmatrix} \sum_{n=1}^{N} \log p(y_n|f_n)$$

$$diag\left[\frac{d^2}{df_1^2}\log p(y_1|f_1),\cdots,\frac{d^2}{df_N^2}\log p(y_N|f_N)\right]=diag\left[\frac{-e^{-y_1f_1}}{(1+e^{-y_1f_1})^2},\cdots,\frac{-e^{-y_Nf_N}}{(1+e^{-y_Nf_N})^2}\right]$$

$$\mathbf{g}_{\psi}(\mathbf{f})$$

$$\mathbf{H}_{\psi}(\mathbf{f})$$

$$\psi(\mathbf{f}) = \log p(\mathbf{f}|\mathcal{D})$$

$$\mathbf{g}_{\psi}(\mathbf{f})$$

$$\frac{d}{d\mathbf{f}}\psi(\mathbf{f}) = -\mathbf{\Sigma}_{f|D}^{-1}(\mathbf{f} - \mathbf{m}_{f|D})$$

$$\mathbf{H}_{\psi}(\mathbf{f})$$

$$\frac{d^2}{d\mathbf{f}^2}\psi(\mathbf{f}) = \frac{d}{d\mathbf{f}}\mathbf{g}_\psi(\mathbf{f}) = -\boldsymbol{\Sigma}_{f|D}^{-1}$$

$$\mathbf{H}_{\psi}(\mathbf{f})$$

$$\mathbf{H}_{\psi}(\mathbf{f}) = \mathbf{W} - \mathbf{K}^{-1} = -\boldsymbol{\Sigma}_{f|D}^{-1},$$

$$\boldsymbol{\Sigma}_{f|D} = (\mathbf{K}^{-1} - \mathbf{W})^{-1}$$

$$\mathbf{g}_\psi(\mathbf{f}) = \mathbf{0}$$

$$\mathbf{f}_{n+1}$$

$$\mathbf{f}_n - \mathbf{H}_\psi^{-1}(\mathbf{f}_n)\, \mathbf{g}_\psi(\mathbf{f}_n) = \mathbf{f}_n + (\mathbf{K}^{-1} - \mathbf{W})^{-1} \left(\mathbf{w} - \mathbf{K}^{-1}\mathbf{f}_n\right)$$

$$\mathbf{f}_n + (\mathbf{K}^{-1} - \mathbf{W})^{-1} \left[-(\mathbf{K}^{-1} - \mathbf{W})\mathbf{f}_n + \mathbf{w} - \mathbf{W}\mathbf{f}_n \right]$$

$$\left(\mathbf{K}^{-1} - \mathbf{W}\right)^{-1} \left(\mathbf{w} - \mathbf{W}\mathbf{f}_n\right)$$

$$\mathbf{f} = \mathbf{m}_{f|D}$$

$$\mathbf{g}_{\psi}(\mathbf{f}) = \frac{d}{d\mathbf{f}}\psi(\mathbf{f}) = \mathbf{w} - \mathbf{K}^{-1}\mathbf{f} = \mathbf{0},$$

$$\mathbf{f} = \mathbf{K}\ \mathbf{w}$$

$$\mathbf{m}_{f|D} = E(\mathbf{f}) = E\left(\mathbf{K}\mathbf{w}\right) = \mathbf{K}E(\mathbf{w}) = \mathbf{K}\mathbf{w}$$

$$p(\mathbf{f}|\mathcal{D}) \approx \mathcal{N}(\mathbf{m}_{f|D}, \boldsymbol{\Sigma}_{f|D})$$

$$\mathbf{f}_* = \mathbf{f}(\mathbf{X}_*)$$

$$p(\mathbf{f}|\mathbf{X}_*, \mathcal{D}) \approx \mathcal{N}(\mathbf{m}_{f*}, \boldsymbol{\Sigma}_{f*})$$

$$\mathbf{m}_{f*}$$

$$\Sigma_{f*}$$

$$p(\mathbf{f}, \mathbf{f}_8)$$

$$\mathbf{m}_{f_*|f} = E(\mathbf{f}_*|\mathbf{f})$$

$$\mathbf{\Sigma}_{f_*|f} = Cov(\mathbf{f}_*|\mathbf{f})$$

$$p(\mathbf{f}_* | \mathbf{X}_*, \mathbf{X}, \mathbf{f})$$

$$\begin{cases} \mathbf{m}_{f_*|f} = \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{f} \\ \boldsymbol{\Sigma}_{f_*|f} = \mathbf{K}_{**} - \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{K}_* \end{cases}$$

$$\mathbf{m}_{f_*} = E_f(\mathbf{f}_*)$$

$$\mathbf{m}_{f_*|f} = \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{f}$$

$$\mathbf{m}_{f_*|f}$$

$$\mathbf{m}_{f*}$$

$$E_f(\mathbf{m}_{f_*|f}) = E_f\left(\mathbf{K}_*^T\mathbf{K}^{-1}\mathbf{f}\right) = \mathbf{K}_*^T\mathbf{K}^{-1}E_f(\mathbf{f}) = \mathbf{K}_*^T\mathbf{K}^{-1}\mathbf{m}_f$$

$$\mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{K} \; \mathbf{w}_{\mathbf{m}_f} = \mathbf{K}_*^T \mathbf{w}_{\mathbf{m}_f}$$

$$\mathbf{m}_f = E_f(\mathbf{f}) = \mathbf{K}\mathbf{w}_{m_f}$$

$$\mathbf{\Sigma}_{m_{f_*|f}} = E_f[(\mathbf{m}_{f_*|f} - \mathbf{m}_{f_*})^2]$$

$$\boldsymbol{\Sigma}_{f|D} = (\mathbf{K}^{-1} - \mathbf{W})^{-1}$$

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

$$\mathbf{\Sigma}_y = \mathbf{A}\mathbf{\Sigma}_x\mathbf{A}^T$$

$$\mathbf{\Sigma}_{m_{f_*|f}} = \mathbf{K}_*^T \mathbf{K}^{-1} \, \mathbf{\Sigma}_{f|D} \, \mathbf{K}^{-1} \mathbf{K}_* = \mathbf{K}_*^T \mathbf{K}^{-1} (\mathbf{K}^{-1} - \mathbf{W})^{-1} \mathbf{K}^{-1} \mathbf{K}_*$$

$\mathbf{f}_*$

$$\mathbf{\Sigma}_{f_*|f} = E[(\mathbf{f}_* - \mathbf{m}_{f_*|f})(\mathbf{f}_* - \mathbf{m}_{f_*|f})^T]$$

$$\Sigma_{f*}$$

$$\boldsymbol{\Sigma}_{f_*|f} + \boldsymbol{\Sigma}_{m_{f_*|f}}$$

$$(\mathbf{K}_{**} - \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{K}_*) + (\mathbf{K}_*^T \mathbf{K}^{-1} (\mathbf{K}^{-1} - \mathbf{W})^{-1} \mathbf{K}^{-1} \mathbf{K}_*)$$

$$\mathbf{K}_{**} - \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{K}_* + \mathbf{K}_*^T \mathbf{K}^{-1} [\mathbf{K} - \mathbf{K}(\mathbf{K} - \mathbf{W}^{-1})^{-1} \mathbf{K}] \mathbf{K}^{-1} \mathbf{K}_*$$

$$\mathbf{K}_{**} - \mathbf{K}_{*}^{T}(\mathbf{K} - \mathbf{W}^{-1})^{-1}\mathbf{K}_{*}$$

$$(\mathbf{A}+\mathbf{B})^{-1}$$

$$p(\mathbf{f}|\mathbf{X}_*, \mathcal{D}) = \mathcal{N}(\mathbf{m}_{f_*}, \mathbf{\Sigma}_{f_*})$$

$$\int \sigma(\mathbf{f}_*)\, p(\mathbf{f}_*|\mathbf{X}_*, \mathcal{D})\, d\mathbf{f}_*$$

$$E_f[\sigma(\mathbf{f}(\mathbf{X}_*))] = \sigma(E_f(\mathbf{f}(\mathbf{X}_*))) = \sigma(\mathbf{m}_{f_*})$$

$$\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$$

K

$\mathbf{K}_*$

$\mathbf{K}_{**}$

$$p(\mathbf{f}|\mathbf{X}, \mathbf{y})$$

$$\mathbf{K} = Cov(\mathbf{X})$$

$\mathbf{m}_{f*}|y$

$$\mathbf{\Sigma}_{f_*|y}$$

$$\mathbf{m}_{f_*|D}$$

$$\Sigma_{f_*|D}$$

$$\mathcal{N}(\mathbf{m}_0, \mathbf{\Sigma}_0)$$

$$\mathcal{N}(\mathbf{m}_1, \boldsymbol{\Sigma}_1)$$

$$\mathbf{m}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{m}_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

$\sigma(\mathbf{m}_{f_*|y})$

α

$\alpha = 0.4$

$\alpha = 0.2$

$$\alpha = 0.03$$

$$C_k, \ (k = 1, \cdots, K)$$

$$\mathbf{y}_k = [y_1^k, \cdots, y_N^k]^T$$

$$\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$$

$$y_n^k = \begin{cases} 1 & \text{if } \mathbf{x}_n \in C_k \\ 0 & \text{if } \mathbf{x}_n \notin C_k \end{cases}$$

$$\mathbf{f}_k = [f_1^k, \cdots, f_N^k]^T$$

$$f_n^k = f^k(\mathbf{x}_n)$$

$$\mathbf{p}_k = [p_1^k, \cdots, p_N^k]^T$$

$$p_n^k \ (n = 1, \cdots, N)$$

$$\mathbf{x}_n \in C_k$$

$f_n^k, \ (k = 1, \cdots, K, \ n = 1, \cdots, N)$

$$p_n^k = p(y_n^k = 1 | f_n^1, \cdots, f_n^K) = \frac{\exp(f_n^k)}{\sum_{l=1}^K \exp(f_n^l)} = \begin{cases} 1 & \text{if } f_n^k = \infty \\ 0 & \text{if } f_n^k = -\infty \end{cases}$$

$$y_n^k = 1$$

$$y_n^l = 0$$

$$l = 1, \cdots, K$$

$$l \neq k$$

$$y_n^1, \cdots, y_n^K$$

$\mathbf{y}_k$

$\mathbf{f}_k$

$\mathbf{p}_k$

KN

$$
\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{bmatrix} = \begin{bmatrix} y_1^1 \\ \vdots \\ y_N^1 \\ \vdots \\ y_1^K \\ \vdots \\ y_N^K \end{bmatrix}, \qquad \mathbf{f} = \begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_K \end{bmatrix} = \begin{bmatrix} f_1^1 \\ \vdots \\ f_N^1 \\ \vdots \\ f_1^K \\ \vdots \\ f_N^K \end{bmatrix}, \qquad \mathbf{p} = \begin{bmatrix} \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_K \end{bmatrix} = \begin{bmatrix} p_1^1 \\ \vdots \\ p_N^1 \\ \vdots \\ p_1^K \\ \vdots \\ p_N^K \end{bmatrix}
$$

$$p(\mathbf{f}|\mathcal{D}) = p(\mathbf{f}|\mathbf{X}, \mathbf{y}) \propto p(\mathbf{y}|\mathbf{f})\ p(\mathbf{f}|\mathbf{X})$$

p_n^k

$$p(\mathbf{y}|\mathbf{f}) = \prod_{n=1}^{N} \prod_{k=1}^{K} (p_n^k)^{y_n^k} = \prod_{n=1}^{N} \prod_{k=1}^{K} \left(\frac{\exp(f_n^k)}{\sum_{h=1}^{K} \exp(f_n^h)} \right)^{y_n^k}$$

$$\{y_n^1, \cdots, y_n^K\}$$

$$y_n^k = 0$$

$$p(\mathbf{f}_k|\mathbf{X}) = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_k)$$

$$cov(f_k(\mathbf{x}_m), f_k(\mathbf{x}_n)) = cov(f_m^k, f_n^k) = k(\mathbf{x}_m, \mathbf{x}_n) = \exp\left(-\frac{1}{a^2}||\mathbf{x}_m - \mathbf{x}_n||^2\right), \qquad (m, n = 1, \cdots, N)$$

$$p(\mathbf{f}|\mathbf{X}) = \mathcal{N}(\mathbf{0}, \mathbf{K})$$

0

$$KN \times KN$$

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{K}_K \end{bmatrix}$$

$$p(\mathbf{f}|\mathcal{D}) = p(\mathbf{f}|\mathbf{X}, \mathbf{y}) \propto p(\mathbf{y}|\mathbf{f})\; p(\mathbf{f}|\mathbf{X}) \propto \prod_{n=1}^{N} \prod_{k=1}^{K} \left(\frac{\exp(f_n^k)}{\sum_{h=1}^{K} \exp(f_n^h)} \right)^{y_n^k} \mathcal{N}(\mathbf{0},\, \mathbf{K})$$

$y_n^k \in \{0,\ 1\}$

$$p(\mathbf{f}|\mathcal{D}) \approx \mathcal{N}(\mathbf{f},\, \mathbf{m}_{f|D},\, \mathbf{\Sigma}_{f|D})$$

$$\psi(\mathbf{f})$$

$$\log p(\mathbf{f}|\mathcal{D}) = \log p(\mathbf{y}|\mathbf{f}) + \log p(\mathbf{f}|\mathbf{X}) = \sum_{n=1}^{N} \sum_{k=1}^{K} y_n^k \left(f_n^k - \log \sum_{h=1}^{K} \exp(f_n^h) \right) + \log \mathcal{N}(\mathbf{0}, \mathbf{K})$$

$$\mathbf{y}^T\mathbf{f} - \sum_{n=1}^{N}\sum_{k=1}^{K}\log\sum_{h=1}^{K}\exp(f_n^h) - \frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{K}| - \frac{1}{2}\mathbf{f}^T\mathbf{K}^{-1}\mathbf{f}$$

$\mathbf{g}_{\psi}$

$$\frac{d}{d\mathbf{f}}\psi(\mathbf{f}) = \frac{d}{d\mathbf{f}}\left(\mathbf{y}^T\mathbf{f} - \sum_{n=1}^{N}\sum_{k=1}^{K}\log\sum_{h=1}^{K}\exp(f_n^h) - \frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{K}| - \frac{1}{2}\mathbf{f}^T\mathbf{K}^{-1}\mathbf{f}\right)$$

$$\mathbf{y} - \mathbf{p} - \mathbf{K}^{-1}\mathbf{f}$$

p

$$\sum_{n=1}^{N}\sum_{k=1}^{K}\frac{d}{d\mathbf{f}}\log\sum_{h=1}^{K}\exp(f_n^h)=\sum_{n=1}^{N}\sum_{k=1}^{K}\begin{bmatrix}0\\ \vdots\\ 0\\ \frac{\exp(f_n^k)}{\sum_{h=1}^{K}\exp(f_n^h)}\\ 0\\ \vdots\\ 0\end{bmatrix}=\sum_{n=1}^{N}\sum_{k=1}^{K}\begin{bmatrix}0\\ \vdots\\ 0\\ p_n^k\\ 0\\ \vdots\\ 0\end{bmatrix}=\mathbf{p}$$

$$p(\mathbf{f}|\mathcal{D}) \approx \mathcal{N}(\mathbf{m}_{f|D}, \mathbf{\Sigma}_{f|D})$$

$$\mathbf{g}_\psi(\mathbf{f})\bigg|_{\mathbf{f}=\mathbf{m}_{f|D}} = \mathbf{y} - \mathbf{p} - \mathbf{K}^{-1}\mathbf{f}\bigg|_{\mathbf{f}=\mathbf{m}_{f|D}} = \mathbf{y} - \mathbf{p}_{m_f} - \mathbf{K}^{-1}\mathbf{m}_{f|D} = \mathbf{0},$$

$$\mathbf{m}_{f|D} = \mathbf{K}(\mathbf{y} - \mathbf{p}_{m_f})$$

$$\mathbf{p}_{m_f} = E_f(\mathbf{p})$$

$$\mathbf{H}_{\psi}(\mathbf{f}) = \frac{d^2}{d\mathbf{f}^2}\psi(\mathbf{f}) = \frac{d}{d\mathbf{f}}\mathbf{g}_{\psi} = \frac{d}{d\mathbf{f}}\Big(\mathbf{y} - \mathbf{K}^{-1}\mathbf{f} - \mathbf{p}\Big) = -\mathbf{K}^{-1} - \mathbf{W} = -\mathbf{\Sigma}_{f|D}^{-1}$$

$$\boldsymbol{\Sigma}_{f|D} = -\mathbf{H}_{\psi}^{-1}(\mathbf{f}) = (\mathbf{K}^{-1} + \mathbf{W})^{-1}$$

$$\mathbf{W} = d\mathbf{p}/d\mathbf{f}$$

$ijkl$

$$(i, j = 1, \cdots, N, \quad k, l = 1, \cdots, K)$$

$$\frac{\partial}{\partial f_j^l} p_i^k$$

$$\frac{\partial}{\partial f_j^l}\left[\frac{\exp(f_i^k)}{\sum_{h=1}^K \exp(f_i^h)}\right] = \frac{\exp(f_i^k)\left(\sum_{h=1}^K \exp(f_i^h)\right)\delta_{kl} - \exp(f_i^k)\exp(f_j^l)}{\left(\sum_{h=1}^K \exp(f_i^h)\right)^2}\delta_{ij}$$

$$\left[\frac{\exp(f_i^k)}{\sum_{h=1}^{K}\exp(f_i^h)}\delta_{kl}-\frac{\exp(f_i^k)\exp(f_j^l)}{\left(\sum_{h=1}^{K}\exp(f_i^h)\right)^2}\right]\delta_{ij}=(p_i^k\delta_{kl}-p_i^k p_j^l)\delta_{ij},$$

$$\mathbf{W} = diag(\mathbf{p}) - \mathbf{P}\mathbf{P}^T$$

$$diag(\mathbf{p})$$

P

$N \times N$

$$diag(\mathbf{p}_k)$$

$$\mathbf{P} = \left[\begin{array}{c} diag(\mathbf{p}_1) \\ \vdots \\ diag(\mathbf{p}_K) \end{array} \right], \quad diag(\mathbf{p}_k) = \left[\begin{array}{cccc} p_1^k & 0 & \cdots & 0 \\ 0 & p_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & p_N^k \end{array} \right], \quad (k = 1, \cdots, K)$$

$$\mathbf{P}\mathbf{P}^T = \begin{bmatrix} diag(\mathbf{p}_1) \\ \vdots \\ diag(\mathbf{p}_K) \end{bmatrix} [diag(\mathbf{p}_1), \cdots, diag(\mathbf{p}_K)] = \begin{bmatrix} diag(\mathbf{p}_1^2) & \cdots & diag(\mathbf{p}_1)diag(\mathbf{p}_K) \\ \vdots & \ddots & \vdots \\ diag(\mathbf{p}_K)diag(\mathbf{p}_1) & \cdots & diag(\mathbf{p}_K^2) \end{bmatrix}$$

$$diag(\mathbf{p}_k)diag(\mathbf{p}_l) = \left[\begin{array}{cccc} p_1^k p_1^l & 0 & \cdots & 0 \\ 0 & p_1^k p_1^l & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & p_N^k p_N^l \end{array} \right]$$

$$\mathbf{g}_\psi = \mathbf{y} - \mathbf{p} - \mathbf{K}^{-1}\mathbf{f}$$

$$\mathbf{H}_{\psi} = -(\mathbf{K}^{-1} + \mathbf{W})$$

$$\mathbf{f}_n - \mathbf{H}_{\psi}^{-1}\mathbf{g}_{\psi} = \mathbf{f}_n + (\mathbf{K}^{-1} + \mathbf{W})^{-1}(\mathbf{y} - \mathbf{K}^{-1}\mathbf{f}_n - \mathbf{p})$$

$$\mathbf{f}_n + (\mathbf{K}^{-1} + \mathbf{W})^{-1}(-(\mathbf{K}^{-1} + \mathbf{W})\mathbf{f}_n + \mathbf{W}\mathbf{f}_n + \mathbf{y} - \mathbf{p})$$

$$(\mathbf{K}^{-1} + \mathbf{W})^{-1}(\mathbf{W}\mathbf{f}_n + \mathbf{y} - \mathbf{p})$$

$$\mathbf{f} = \mathbf{K}(\mathbf{y} - \mathbf{p})$$

$$E_f(\mathbf{m}_{f_*|f}) = E_f(\mathbf{K}_*^T\mathbf{K}^{-1}\mathbf{f}) = \mathbf{K}_*^T\mathbf{K}^{-1}E_f(\mathbf{f})$$

$$\mathbf{K}_*^T \mathbf{K}^{-1} \left[\mathbf{K}(\mathbf{y} - \mathbf{p}_{m_f}) \right] = \mathbf{K}_*^T (\mathbf{y} - \mathbf{p}_{m_f})$$

$$\mathbf{m}_{f_*^k} = (\mathbf{K}_*^k)^T(\mathbf{y}^k - \mathbf{p}^k), \qquad (k = 1, \cdots, K)$$

$$\mathbf{\Sigma}_{f|D} = (\mathbf{K}^{-1} + \mathbf{W})^{-1}$$

$$\mathbf{\Sigma}_{m_{f_*|f}} = \mathbf{K}_*^T \mathbf{K}^{-1} \, \mathbf{\Sigma}_{f|D} \, \mathbf{K}^{-1} \mathbf{K}_* = \mathbf{K}_*^T \mathbf{K}^{-1} (\mathbf{K}^{-1} + \mathbf{W})^{-1} \mathbf{K}^{-1} \mathbf{K}_*$$

$$\mathbf{W} = diag(\mathbf{p}) - \mathbf{P}\mathbf{P}^T$$

$$\mathbf{\Sigma}_{f_*} = \mathbf{K}_{**} - \mathbf{K}_*^T(\mathbf{K} + \mathbf{W}^{-1})^{-1}\mathbf{K}_*$$

$$p_*^k = \frac{\exp(m_{f_*}^k)}{\sum_{l=1}^K \exp(m_{f_*}^l)}, \qquad (k = 1, \cdots, K)$$

$$p_*^k = \max\{p_*^1, \cdots, p_*^K\}$$

$$p(\mathbf{f}_* | \mathbf{X}, \mathbf{y}, \mathbf{X}_*)$$

$$30/600 = 0.045$$

$$\begin{bmatrix} 197 & 3 & 0 \\ 0 & 195 & 5 \\ 1 & 21 & 178 \end{bmatrix}$$

$$13/400 = 0.0325$$

$$\begin{bmatrix} 193 & 7 \\ 6 & 194 \end{bmatrix}$$

D

$d < D$

$$d \ll D$$

n_k

$$d_B(C_i, C_j) = \frac{1}{4}(\mathbf{m}_i - \mathbf{m}_j)^T \left[\frac{\mathbf{\Sigma}_i + \mathbf{\Sigma}_j}{2}\right]^{-1} (\mathbf{m}_i - \mathbf{m}_j) + \log\left[\frac{\left|\frac{\mathbf{\Sigma}_i + \mathbf{\Sigma}_j}{2}\right|}{(|\mathbf{\Sigma}_i| \ \ |\mathbf{\Sigma}_j|)^{1/2}}\right]$$

d_B

$$C_i \cup C_j = C_k$$

$$\mathbf{m}_k = \frac{1}{n_i + n_j}[n_i \mathbf{m}_i + n_j \mathbf{m}_j]$$

$$\mathbf{\Sigma}_k = \frac{1}{n_i + n_j}[n_i(\Sigma_i + (\mathbf{m}_i - \mathbf{m}_k)(\mathbf{m}_i - \mathbf{m}_k)^T) + n_j(\Sigma_j + (\mathbf{m}_j - \mathbf{m}_k)(\mathbf{m}_j - \mathbf{m}_k)^T)]$$

$K - 2$

$\mathbf{S}_B$

λ_i

$\mathbf{v}_i$

$\mathbf{v}_1$

$$y_n = \mathbf{x}_n^T \mathbf{v} \qquad (n = 1, \cdots, N)$$

G_l

G_r

$D_l(\mathbf{x})$

$$D_r(\mathbf{x})$$

M

$$\mathbf{x} \in \begin{cases} G_l & if \quad D_l(\mathbf{x}) > D_r(\mathbf{x}) \\ G_r & if \quad D_r(\mathbf{x}) < D_l(\mathbf{x}) \end{cases}$$

$$N = 4$$

$$\mathbf{x}_n \in \mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N\}]$$

$$y_n \in \mathbf{y} = [y_1, \cdots, y_n]^T$$

$\mathbf{m}_1, \cdots, \mathbf{m}_K$

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} P_{nk} ||\mathbf{x}_n - \mathbf{m}_k||^2$$

$$P_{nk} = \begin{cases} 1 & \text{if } k = \arg\min_l ||\mathbf{x}_n - \mathbf{m}_l|| \\ 0 & \text{otherwise} \end{cases}$$

J

$$\frac{d}{d\mathbf{m}_k}J = 2\sum_{n=1}^{N} P_{nk}||\mathbf{x}_n - \mathbf{m}_k|| = \mathbf{0}$$

$$\mathbf{m}_k = \frac{\sum_{n=1}^{N} P_{nk}\mathbf{x}_n}{\sum_{n=1}^{N} P_{nk}} = \frac{1}{N_k}\sum_{n=1}^{N} P_{nk}\mathbf{x}_n$$

$$N_k = \sum_{n=1}^{N} P_{nk}$$

$$\mathbf{m}_1^{(0)},\ \mathbf{m}_2^{(0)}, ..., \mathbf{m}_K^{(0)}$$

$$l = 1$$

$\mathbf{x} \in \mathbf{X}$

$$||\mathbf{x} - \mathbf{m}_k^{(l)}||^2 = \min_{1 \leq l \leq K} ||\mathbf{x} - \mathbf{m}_l^{(l)}||^2,$$

$$\mathbf{x} \in C_k^{(l)}$$

$$C_k^{(l)}$$

$$\mathbf{m}_k^{(l)}$$

$$\mathbf{m}_k^{(l+1)}$$

$$\mathbf{m}_k^{l+1}$$

$$\mathbf{m}_k^{(l+1)} = \frac{1}{N_k} \sum_{\mathbf{x} \in C_k} \mathbf{x}, \qquad (k = 1, \cdots, K)$$

$$\mathbf{m}_k^{(l+1)} = \mathbf{m}_k^{(l)} \quad (k = 1, \cdots, K)$$

$$l \leftarrow l + 1$$

$$tr(\mathbf{S}_T^{-1}\mathbf{S}_B)$$

$$C = 4$$

$$K = C - 1$$

$$K = C$$

$$K = C + 1$$

$$K = 3,\ 4,\ 5$$

$$tr(\mathbf{\Sigma}_k) \ \ (k = 1, \cdots, K)$$

	K=C-1=3			K=C=4				K=C+1=5				
Separability	1.76			2.56				2.58				
Intra-cluster distance	9.1	44.3	11.8	10.8	12.7	11.1	9.1	10.8	9.1	11.1	8.9	9.4

K=C-1=3		
	1	2
2	10.9	
3	21.9	184.7

K=C=4			
	1	2	3
2	4.0		
3	5.1	1.6	
4	2.4	2.3	4.4

K=C+1=5				
	1	2	3	4
2	4.0			
3	5.1	1.6		
4	7.0	4.3	0.3	
5	17.3	15.5	8.9	11.1

$$K = 3 < C = 4$$

$$K = 5 > C = 4$$

$K = 3$

$$K = 5$$

$$\begin{bmatrix} 1 & 27 & 3 & 166 & 0 & 0 & 3 & 24 & 0 & 0 \\ 0 & 1 & 13 & 0 & 210 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 16 & 0 & 3 & 180 & 9 & 8 & 5 & 0 \\ 1 & 0 & 88 & 1 & 0 & 0 & 5 & 1 & 128 & 0 \\ 4 & 1 & 52 & 0 & 26 & 0 & 0 & 6 & 0 & 135 \\ 1 & 1 & 10 & 0 & 0 & 0 & 0 & 167 & 43 & 2 \\ 0 & 161 & 5 & 14 & 13 & 2 & 0 & 29 & 0 & 0 \\ 4 & 0 & 70 & 0 & 5 & 4 & 137 & 3 & 1 & 0 \\ 4 & 1 & 69 & 2 & 1 & 3 & 2 & 31 & 110 & 1 \\ 92 & 0 & 101 & 1 & 1 & 1 & 5 & 0 & 2 & 21 \end{bmatrix}$$

$$p(\mathbf{x}) = \sum_{k=1}^{K} P_k \mathcal{N}(\mathbf{x}; \mathbf{m}_k, \boldsymbol{\Sigma}_k) = \sum_{k=1}^{K} P_k \left[\frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_k|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{m}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \mathbf{m}_k) \right) \right]$$

$$\mathcal{N}(\mathbf{x}; \mathbf{m}_k, \mathbf{\Sigma}_k)$$

$$\int_{-\infty}^{\infty} p(\mathbf{x})\, d\mathbf{x} = \sum_{k=1}^{K} P_k \int_{-\infty}^{\infty} \mathcal{N}(\mathbf{x}; \mathbf{m}_k, \boldsymbol{\Sigma}_k)\, d\mathbf{x} = \sum_{k=1}^{K} P_k = 1$$

$$\mathcal{N}(\mathbf{x}; \mathbf{m}_k, \mathbf{\Sigma}_k), \ (k = 1, \cdots, K)$$

$$\theta = \{P_k, \mathbf{m}_k, \mathbf{\Sigma}_k,\ (k = 1, \cdots, K)\}$$

$$P_{nk} = P(\mathbf{x}_n \in C_k)$$

$$P_{nk} = \max_l P_{nl}$$

$$\mathcal{N}(\mathbf{x}, \mathbf{m}_k, \mathbf{\Sigma}_k)$$

$$\mathbf{z} = [z_1, \cdots, z_K]^T$$

$$z_k \in \{0,\, 1\}$$

$$z_k = 1$$

$$\sum_{k=1}^{K} z_k = 1$$

$$C_k \quad (k = 1, \cdots, K)$$

$$P(z_k = 1) = P_k$$

$$\mathbf{x} \in C_k, \ (k = 1, \cdots, K)$$

$$\sum_{k=1}^{K} P_k = 1$$

$$p(\mathbf{x}|z_k = 1, \theta) = \mathcal{N}(\mathbf{x}; \mathbf{m}_k, \mathbf{\Sigma}_k)$$

$$p(\mathbf{x}, z_k = 1|\theta) = p(\mathbf{x}|z_k = 1, \theta)\ P(z_k = 1) = P_k \mathcal{N}(\mathbf{x}; \mathbf{m}_k, \mathbf{\Sigma}_k)$$

$$p(\mathbf{x}|\theta) = \sum_{k=1}^{K} p(\mathbf{x}, z_k = 1|\theta) = \sum_{k=1}^{K} p(\mathbf{x}|z_k = 1, \theta)\ P(z_k = 1) = \sum_{k=1}^{K} P_k \mathcal{N}(\mathbf{x}; \mathbf{m}_k, \mathbf{\Sigma}_k)$$

$$p(\mathbf{z}|\theta)$$

$$\prod_{k=1}^{K} P_k^{z_k}$$

$$p(\mathbf{x}|\mathbf{z}, \theta)$$

$$\prod_{k=1}^{K} \mathcal{N}(\mathbf{x}, \mathbf{m}_k, \mathbf{\Sigma}_k)^{z_k}$$

$$p(\mathbf{x}, \mathbf{z}|\theta)$$

$$p(\mathbf{z})\ p(\mathbf{x}|\mathbf{z},\theta) = \prod_{k=1}^{K} \left(P_k \mathcal{N}(\mathbf{x}, \mathbf{m}_k, \boldsymbol{\Sigma}_k)\right)^{z_k}$$

$$\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_N]$$

$$\mathbf{z}_n = [z_{n1}, \cdots, z_{nK}]^T$$

$$z_{nk} = 1$$

$$z_{nl} = 0$$

$$\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_N]$$

Y

Z

$$p(\mathbf{x}_n, \mathbf{z}_n|\theta) = \prod_{k=1}^{K} \left(P_k \mathcal{N}(\mathbf{x}_n, \mathbf{m}_k, \mathbf{\Sigma}_k)\right)^{z_{nk}}, \qquad (n = 1, \cdots, N)$$

θ

$$L(\theta|\mathbf{X}, \mathbf{Z})$$

$$p(\mathbf{X}, \mathbf{Z}|\theta) = p([\mathbf{x}_1, \cdots, \mathbf{x}_N], [\mathbf{z}_1, \cdots, \mathbf{z}_N] \Big| \mathbf{m}_k, \mathbf{\Sigma}_k, P_k (k = 1, \cdots, K))$$

$$\prod_{n=1}^{N} p(\mathbf{x}_n, \mathbf{z}_n | \theta) = \prod_{n=1}^{N} \prod_{k=1}^{K} \left(P_k \mathcal{N}(\mathbf{x}_n, \mathbf{m}_k, \mathbf{\Sigma}_k)\right)^{z_{nk}}$$

$$\log L(\theta|\mathbf{X}, \mathbf{Z})$$

$$\log p(\mathbf{X}, \mathbf{Z}|\theta) = \log \left[\prod_{n=1}^{N} \prod_{k=1}^{K} \left(P_k \mathcal{N}(\mathbf{x}_n, \mathbf{m}_k, \mathbf{\Sigma}_k) \right)^{z_{nk}} \right]$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\left[\log P_k + \log \mathcal{N}(\mathbf{x}_n, \mathbf{m}_k, \mathbf{\Sigma}_k)\right]$$

$L(\theta)$

$$\log\ L(\theta)$$

$$P_{nk} = P(z_{nk} = 1|\mathbf{x}_n, \theta) = \frac{p(\mathbf{x}_n, z_{nk} = 1|\theta)}{p(\mathbf{x}_n|\theta)} = \frac{P_k\,\mathcal{N}(\mathbf{x}_n; \mathbf{m}_k \boldsymbol{\Sigma}_k)}{\sum_{l=1}^{K} P_l\,\mathcal{N}(\mathbf{x}_n; \mathbf{m}_l \boldsymbol{\Sigma}_l)} \qquad (n = 1, \cdots, N;\; k = 1, \cdots, K)$$

$$P(z_k = 1) = P_k$$

$$\sum_{k=1}^{K} P(z_k = 1) = \sum_{k=1}^{K} P_k = 1, \qquad \sum_{k=1}^{K} P(z_{nk} = 1|\mathbf{x}_n, \theta) = \sum_{k=1}^{K} P_{nk} = 1$$

P_{nk}

s_{nk}

$$0 < P_{nk} < 1$$

$$P_{nk} = 1$$

$$P_{nl} = 0$$

$$E_{\mathbf{Z}}\left(\log L(\theta|\mathbf{X},\mathbf{Z})\right)$$

$$E_{\mathbf{Z}}\left[\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\left[\log P_k + \log \mathcal{N}(\mathbf{x}_n, \mathbf{m}_k, \boldsymbol{\Sigma}_k)\right]\right]$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K}E(z_{nk})\left[\log P_k + \log\mathcal{N}(\mathbf{x}_n, \mathbf{m}_k, \mathbf{\Sigma}_k)\right]$$

$$\sum_{n=1}^{N} \sum_{k=1}^{K} P_{nk} \left[\log P_k + \log \mathcal{N}(\mathbf{x}_n, \mathbf{m}_k, \mathbf{\Sigma}_k) \right]$$

$$E(z_{nk}) = 1\ P(z_{nk} = 1|\mathbf{x}_n) + 0\ P(z_{nk} = 0|\mathbf{x}_n) = P(z_{nk} = 1|\mathbf{x}_n) = P_{nk}$$

$$\theta = \{P_k,\ \mathbf{m}_k\ (k = 1, \cdots, K)\}$$

$$\sum_{k=1}^{K} P_k = 1$$

$$L(\theta,\ \lambda) = \sum_{n=1}^{N} \sum_{k=1}^{K} P_{nk} \left[\log P_k + \log \mathcal{N}(\mathbf{x}_n, \mathbf{m}_k, \mathbf{\Sigma}_k) \right] + \lambda \left(\sum_{k=1}^{K} P_k - 1 \right)$$

$$\frac{\partial}{\partial P_k} L(\theta,\ \lambda)$$

$$\frac{\partial}{\partial P_k}\left[\sum_{n=1}^{N}\sum_{k=1}^{K}P_{nk}\left[\log P_k+\log\mathcal{N}(\mathbf{x}_n,\mathbf{m}_k,\mathbf{\Sigma}_k)\right]+\lambda\left(\sum_{k=1}^{K}P_k-1\right)\right]$$

$$\sum_{n=1}^{N} P_{nk} \frac{1}{P_k} + \lambda = 0$$

$$\sum_{n=1}^{N} P_{nk} + P_k\lambda = N_k + P_k\lambda = 0$$

$$N_k = \sum_{n=1}^{N} P_{nk}$$

$$\sum_{k=1}^{K} N_k = \sum_{k=1}^{K} \left(\sum_{n=1}^{N} P_{nk} \right) = \sum_{n=1}^{N} \left(\sum_{k=1}^{K} P_{nk} \right) = \sum_{n=1}^{N} 1 = N$$

$$\sum_{k=1}^{K} (N_k + P_k \lambda) = \sum_{k=1}^{K} N_k + \lambda \left(\sum_{k=1}^{K} P_k \right) = N + \lambda = 0$$

$$\lambda = -N$$

$$p(z_k = 1) = P_k = \frac{N_k}{N} = \frac{1}{N}\sum_{n=1}^{N} P_{nk}$$

$$\frac{\partial}{\partial \mathbf{m}_k} E_{\mathbf{z}}(\log L(\theta|\mathbf{X}, \mathbf{Z}))$$

$$\frac{\partial}{\partial \mathbf{m}_k} \left[\sum_{n=1}^{N} \sum_{k=1}^{K} P_{nk} \left[\log P_k + \log \mathcal{N}(\mathbf{x}_n, \mathbf{m}_k, \mathbf{\Sigma}_k) \right] \right]$$

$$\sum_{n=1}^{N} P_{nk} \frac{\partial}{\partial \mathbf{m}_k} \log \mathcal{N}(\mathbf{x}_n, \mathbf{m}_k, \mathbf{\Sigma}_k)$$

$$\sum_{n=1}^{N} P_{nk} \frac{\partial}{\partial \mathbf{m}_k} \left[-\frac{1}{2} (\mathbf{x}_n - \mathbf{m}_k)^T \mathbf{\Sigma}_k^{-1} (\mathbf{x}_n - \mathbf{m}_k) \right]$$

$$\frac{1}{2}\sum_{n=1}^{N} P_{nk}\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n - \mathbf{m}_k) = \mathbf{0}$$

$$(2\pi)^{-d/2}\,|\boldsymbol{\Sigma}_k|^{-1/2}$$

$$2\boldsymbol{\Sigma}_k$$

$$\sum_{n=1}^{N} P_{nk}(\mathbf{x}_n - \mathbf{m}_k) = \sum_{n=1}^{N} P_{nk}\mathbf{x}_n - \sum_{n=1}^{N} P_{nk}\mathbf{m}_k = \sum_{n=1}^{N} P_{nk}\mathbf{x}_n - N_k\mathbf{m}_k = \mathbf{0}$$

$$\mathbf{m}_k = \frac{1}{N_k} \sum_{n=1}^{N} P_{nk} \mathbf{x}_n$$

$$\frac{\partial}{\partial \mathbf{\Sigma}_k} \left[\sum_{n=1}^{N} \sum_{k=1}^{K} P_{nk} \left[\log P_k + \log \mathcal{N}(\mathbf{x}_n, \mathbf{m}_k, \mathbf{\Sigma}_k) \right] \right]$$

$$\sum_{n=1}^{N} P_{nk} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \log \mathcal{N}(\mathbf{x}_n, \mathbf{m}_k, \boldsymbol{\Sigma}_k)$$

$$-\frac{1}{2}\sum_{n=1}^{N} P_{nk}\left[\frac{\partial}{\partial \mathbf{\Sigma}_k}\log|\mathbf{\Sigma}_k| + \frac{\partial}{\partial \mathbf{\Sigma}_k}(\mathbf{x}_n - \mathbf{m}_k)^T \mathbf{\Sigma}_k^{-1}(\mathbf{x}_n - \mathbf{m}_k)\right]$$

$$-\frac{1}{2}\sum_{n=1}^{N} P_{nk}\left[\mathbf{\Sigma}_k^{-1} - \mathbf{\Sigma}_k^{-1}(\mathbf{x}_n - \mathbf{m}_k)(\mathbf{x}_n - \mathbf{m}_k)^T\mathbf{\Sigma}_k^{-1}\right] = \mathbf{0}$$

$$\frac{d}{d\mathbf{A}} \log |\mathbf{A}| = (\mathbf{A}^{-1})^T, \qquad \frac{d}{d\mathbf{A}} \left(\mathbf{a}^T \mathbf{A}^{-1} \mathbf{b} \right) = -(\mathbf{A}^{-1})^T \mathbf{a} \mathbf{b}^T (\mathbf{A}^{-1})^T$$

$$\sum_{n=1}^{N} P_{nk} \left(\mathbf{\Sigma}_k - (\mathbf{x}_n - \mathbf{m}_k)(\mathbf{x}_n - \mathbf{m}_k)^T \right) = \mathbf{0}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} P_{nk}(\mathbf{x}_n - \mathbf{m}_k)(\mathbf{x}_n - \mathbf{m}_k)^T$$

$$\theta = \{P_k, \mathbf{m}_k, \boldsymbol{\Sigma}_k, (k = 1, \cdots, K)\}$$

$$P_{nk} = P(r_k = 1|\mathbf{x}_n) = \frac{P_k\,\mathcal{N}(\mathbf{x}_n;\mathbf{m}_k,\boldsymbol{\Sigma}_k)}{\sum_{l=1}^{K} P_l\,\mathcal{N}(\mathbf{x}_n;\mathbf{m}_l,\boldsymbol{\Sigma}_l)}, \qquad N_k = \sum_{n=1}^{N} P_{nk}$$

P_k

$$\frac{N_k}{N}$$

$\mathbf{m}_k$

$$\frac{1}{N_k} \sum_{n=1}^{N} P_{nk} \mathbf{x}_n$$

$$\mathbf{\Sigma}_k$$

$$\frac{1}{N_k}\sum_{n=1}^{N} P_{nk}(\mathbf{x}_n - \mathbf{m}_k)(\mathbf{x}_n - \mathbf{m}_k)^T$$

C_l

$$p_{nl} = P(z_{nl} = 1|\mathbf{x}_n, \theta)$$

$$P_{nk} = \max_l p_{nl}$$

$$\mathbf{\Sigma}_k = \varepsilon \mathbf{I}$$

$\mathcal{E}$

$$p(\mathbf{x}|z_k = 1, \theta) = \mathcal{N}(\mathbf{x}|\mathbf{m}_k, \varepsilon\mathbf{I}) = \frac{1}{(2\pi)^{d/2}\varepsilon^{1/2}} \exp\left(-\frac{1}{2\varepsilon}||\mathbf{x} - \mathbf{m}_k||^2\right)$$

$$\mathbf{x}_n \in \mathbf{X}$$

$$P_{nk} = P(z_k = 1|\mathbf{x}_n, \theta) = \frac{P_k \mathcal{N}(\mathbf{x}_n; \mathbf{m}_l \boldsymbol{\Sigma}_l)}{\sum_{l=1}^{K} P_l \mathcal{N}(\mathbf{x}_n; \mathbf{m}_l \boldsymbol{\Sigma}_l)} = \frac{P_k \exp(-||\mathbf{x}_n - \mathbf{m}_k||^2/2\varepsilon)}{\sum_{l=1}^{K} P_l \exp(-||\mathbf{x}_n - \mathbf{m}_l||^2/2\varepsilon)}$$

$\varepsilon \to 0$

$$||\mathbf{x} - \mathbf{m}_k||$$

$$P_{nk} = 0$$

$$\lim_{\varepsilon \to 0} P_{nk} = \lim_{\varepsilon \to 0} \frac{P_k \exp(-||\mathbf{x}_n - \mathbf{m}_k||^2/2\varepsilon)}{\sum_{l=1}^{K} P_l \exp(-||\mathbf{x}_n - \mathbf{m}_l||^2/2\varepsilon)} = \begin{cases} 1 & \text{if } ||\mathbf{x}_n - \mathbf{m}_\mathbf{k}|| = \min_l ||\mathbf{x}_n - \mathbf{m}_\mathbf{l}|| \\ 0 & \text{otherwise} \end{cases}$$

$$P_{nk} \in \{0,\ 1\}$$

m

$$\Sigma$$

$$P_{nk} = p(z_{nk} = 1|\mathbf{x}_n, \theta)$$

$$\phi_{nk} = P(y' = k|\mathbf{x}_n)$$

K-means

0	0	50
50	0	0
18	32	0

EM

0	0	50
45	5	0
0	50	0

1

μ

$1 - \mu$

$$\mathcal{B}(x|\mu) = \mu^x (1-\mu)^{1-x} = \left\{ \begin{array}{cl} \mu & \text{if } x = 1 \\ 1-\mu & \text{if } x = 0 \end{array} \right.$$

$$E(x)$$

$$1\ P(x=1)+0\ P(x=0)=1\ \mu+0\ (1-\mu)=\mu$$

$$Var(x)$$

$$E[(x - E(x))^2] = E(x^2) - E(x)^2$$

$$1^2\ P(x=1)+0^2\ P(x=0)-\mu^2=\mu-\mu^2=\mu(1-\mu)$$

$E(\mathbf{x})$

$$\mathbf{m} = [\mu_1, \cdots, \mu_N]^T$$

$$Cov(\mathbf{x})$$

$$\mathbf{\Sigma} = diag(\mu_i(1-\mu_i)) = \begin{bmatrix} \mu_1(1-\mu_1) & & 0 \\ & \ddots & \\ 0 & & \mu_d(1-\mu_d) \end{bmatrix}$$

$$\{\mu_1, \cdots, \mu_N\}$$

$$\mathcal{B}(\mathbf{x}|\mathbf{m}) = \prod_{i=1}^{d} \mathcal{B}(x_i|\mu_i) = \prod_{i=1}^{d} \mu_i^{x_i}(1-\mu_i)^{1-x_i}$$

$$\log \mathcal{B}(\mathbf{x}|\mathbf{m}) = \log \left(\prod_{i=1}^{d} \mathcal{B}(x_i|\mu_i) \right) = \sum_{i=1}^{d} \left[x_i \log \mu_i + (1 - x_i) \log(1 - \mu_i) \right]$$

$$p(\mathbf{x}|\mathbf{m}_k, P_k, (k = 1, \cdots, K)) = p(\mathbf{x}|\theta) = \sum_{k=1}^{K} P_k \mathcal{B}(\mathbf{x}, \mathbf{m}_k) = \sum_{k=1}^{K} P_k \prod_{i=1}^{d} \mu_{ki}^{x_i} (1 - \mu_{ki})^{1-x_i}$$

$$\theta = \{\mathbf{m}_k, P_k, (k = 1, \cdots, K)\}$$

$$\mathbf{m}_k = E_k(\mathbf{x})$$

$\mathcal{B}(\mathbf{x}|\mathbf{m}_k)$

$$\mathbf{m} = E(\mathbf{x}) = \sum_{k=1}^{K} P_k E_k(\mathbf{x}) = \sum_{k=1}^{K} P_k \mathbf{m}_k$$

$$z_k \in \{0,\ 1\}$$

$$\prod_{k=1}^{K} \mathcal{B}(\mathbf{x}, \mathbf{m}_k)^{z_k}$$

$$p(\mathbf{z}|\theta)\ p(\mathbf{x}|\mathbf{z},\theta) = \prod_{k=1}^{K} \left(P_k\ \mathcal{B}(\mathbf{x},\mathbf{m}_k)\right)^{z_k}$$

$$\theta = \{P_k, \mathbf{m}_k,\ (k = 1, \cdots, K)\}$$

$$L(\theta|\mathbf{X},\mathbf{Z}) = p(\mathbf{X},\mathbf{Z}|\theta)$$

$$p([\mathbf{x}_1, \cdots, \mathbf{x}_N], [\mathbf{z}_1, \cdots, \mathbf{z}_N] | \mathbf{m}_k, P_k (k = 1, \cdots, K))$$

$$\prod_{n=1}^{N} p(\mathbf{x}_n, \mathbf{z}_n|\theta) = \prod_{n=1}^{N} \prod_{k=1}^{K} \left(P_k \mathcal{B}(\mathbf{x}_n, \mathbf{m}_k)\right)^{z_{nk}}$$

$$\log\ L(\theta|\mathbf{X},\mathbf{Z})$$

$$\log p(\mathbf{X}, \mathbf{Z}|\theta) = \log \prod_{n=1}^{N} \prod_{k=1}^{K} \left(P_k \mathcal{B}(\mathbf{x}_n, \mathbf{m}_k)\right)^{z_{nk}}$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\left[\log P_k + \log \mathcal{B}(\mathbf{x}_n, \mathbf{m}_k)\right]$$

$$P_{nk} = P(z_{nk} = 1|\mathbf{x}_n, \theta) = \frac{p(\mathbf{x}_n, z_{nk} = 1|\theta)}{p(\mathbf{x}_n)|\theta} = \frac{P_k\, \mathcal{B}(\mathbf{x}_n; \mathbf{m}_k)}{\sum_{l=1}^{K} P_l\, \mathcal{B}(\mathbf{x}_n; \mathbf{m}_l)} \qquad (n = 1, \cdots, N; \quad k = 1, \cdots, K)$$

z_{nk}

$$E_{\mathbf{Z}}\left(\log L(\theta|\mathbf{X},\mathbf{Z})\right)$$

$$E_{\mathbf{Z}} \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \left[\log P_k + \log \mathcal{B}(\mathbf{x}_n, \mathbf{m}_k) \right]$$

$$\sum_{n=1}^{N} \sum_{k=1}^{K} E(z_{nk}) \left[\log P_k + \log \prod_{i=1}^{d} \mu_{ki}^{x_{ni}} (1 - \mu_{ki})^{1 - x_{ni}} \right]$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K} P_{nk}\left[\log P_k + \sum_{i=1}^{d}\left[x_{ni}\log \mu_{ki} + (1-x_{ni})\log(1-\mu_{ki})\right]\right]$$

$$P_k = \frac{N_k}{N} = \frac{1}{N}\sum_{n=1}^{N} P_{nk}$$

$$\frac{\partial}{\partial \mathbf{m}_k} E_{\mathbf{Z}} \left(\log p(\mathbf{X}, \mathbf{Z} | \theta) \right)$$

$$\frac{\partial}{\partial \mathbf{m}_k} \sum_{n=1}^{N} \sum_{k=1}^{K} P_{nk} \left[\log P_k + \sum_{i=1}^{d} \left[x_{ni} \log \mu_{ki} + (1 - x_{ni}) \log(1 - \mu_{ki}) \right] \right]$$

$$\sum_{n=1}^{N} P_{nk} \frac{\partial}{\partial \mathbf{m}_k} \sum_{i=1}^{d} \left[x_{ni} \log \mu_{ki} + (1 - x_{ni}) \log(1 - \mu_{ki}) \right] = \mathbf{0}$$

$$\sum_{n=1}^{N} P_{nk} \frac{d}{d\mu_{ki}} \left[x_{ni} \log \mu_{ki} + (1 - x_{ni}) \log(1 - \mu_{ki}) \right]$$

$$\sum_{n=1}^{N} P_{nk} \left(\frac{x_{ni}}{\mu_{ki}} - \frac{1 - x_{ni}}{1 - \mu_{ki}} \right) = 0$$

$$(1-\mu_{ki})\sum_{n=1}^{N} P_{nk}x_{ni} = \mu_{ki}\sum_{n=1}^{N} P_{nk}(1-x_{ni}) = \mu_{ki}N_k - \mu_{ki}\sum_{n=1}^{N} P_{nk}x_{ni}$$

μ_{ki}

$$\mu_{ki} = \frac{1}{N_k} \sum_{n=1}^{N} P_{nk} x_{ni} \qquad (i = 1, \cdots, d)$$

$K = 10$

$$K = 12$$

$$y = f(\mathbf{x}) + r = \mathbf{x}^T \mathbf{w} + r$$

$$\{C_0,\, C_1\}$$

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}\begin{cases} < 0 \\ > 0 \end{cases},$$

$$\mathbf{x} \in \left\{ \begin{array}{l} C_0 \\ C_1 \end{array} \right.$$

$$\mathcal{D} = \{(\mathbf{x}_n,\, y_n) | n = 1, \cdots, N\} = \{\mathbf{X}, \mathbf{y}\}$$

$$y_i = 1$$

$$\mathbf{x}_i \in C_1$$

$$y_i = -1$$

$\mathbf{x}_i \in C_0$

$$\varepsilon(\mathbf{w}) = ||\mathbf{r}||^2$$

$$p(\mathbf{x} \in C_1|\mathbf{w})$$

$$p(y = 1|\mathbf{x}, \mathbf{w})$$

$$p(\mathbf{x} \in C_0|\mathbf{w})$$

$$p(y = -1|\mathbf{x}, \mathbf{w}) = 1 - p(y = 1|\mathbf{x}, \mathbf{w})$$

$$\sigma(z) =$$

$$\frac{1}{1+e^{-z}} = \frac{e^z}{1+e^z}$$

$$\phi(z) =$$

$$\int_{-\infty}^{z} \mathcal{N}(u|0,1)\, du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{1}{2}u^2\right)\, du$$

$$\sigma(z) = \begin{cases} 0 & z = -\infty \\ 1 & z = \infty \end{cases}$$

$$\sigma(-z) = 1 - \sigma(z)$$

$$f = \mathbf{x}^T \mathbf{w}$$

$$(-\infty,\ \infty)$$

$$\sigma(f(\mathbf{x}))$$

$(0,\ 1)$

$$\sigma(\mathbf{x}^T\mathbf{w}) = \sigma(f)$$

$$p(y = -1|\mathbf{x}, \mathbf{w})$$

$$1 - p(y = 1|\mathbf{x}, \mathbf{w}) = 1 - \sigma(f) = \sigma(-f)$$

$$p(y|\mathbf{x}, \mathbf{w}) = \sigma(y\ \mathbf{x}^T \mathbf{w}) = \sigma(yf),$$

$$p(y|\mathbf{x}, \mathbf{w}) = \phi(y\mathbf{x}^T\mathbf{w}) = \phi(yf)$$

$$\sigma(z)$$

$$\phi(z)$$

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_w)$$

$$\boldsymbol{\Sigma}_w = \mathbf{I}$$

$$\mathcal{L}(\mathbf{w}|\mathcal{D}) = p(\mathcal{D}|\mathbf{w}) = p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \prod_{n=1}^{N} p(y_n|\mathbf{x}_n, \mathbf{w}) = \prod_{n=1}^{N} \sigma(y_i f_i) = \prod_{n=1}^{N} \sigma(y_n \mathbf{x}_n^T \mathbf{w})$$

$$\mathbf{y} = \pm 1$$

$$p(\mathbf{w})$$

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w})$$

$$p(\mathbf{w}|\mathcal{D})$$

$$p(\mathbf{w}|\mathbf{X},\mathbf{y}) = \frac{p(\mathbf{y},\mathbf{w}|\mathbf{X})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\mathbf{X},\mathbf{w})\ p(\mathbf{w})}{p(\mathbf{y}|\mathbf{X})} \propto p(\mathbf{y}|\mathbf{X},\mathbf{w})\ p(\mathbf{w}|\mathbf{X})$$

$$\prod_{n=1}^{N} p(y_n|\mathbf{x}_n, \mathbf{w})\; \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_w) = \prod_{n=1}^{N} \sigma(y_n f_n)\; \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_w|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{w}^T \boldsymbol{\Sigma}_w^{-1}\mathbf{w}\right)$$

$$p(\mathbf{y}|\mathbf{X}) = p(\mathcal{D})$$

$$\psi(\mathbf{w})$$

$$\psi(\mathbf{w}) = \log\ p(\mathbf{w}|\mathcal{D}) = \sum_{n=1}^{N} \log \sigma(y_n\, f_n) - \frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{\Sigma}_w| - \frac{1}{2}\mathbf{w}^T\mathbf{\Sigma}_w^{-1}\mathbf{w}$$

$$p(\mathbf{w}|\mathcal{D})$$

$$\mathbf{g}_{\psi}(\mathbf{w})$$

$$\frac{d}{d\mathbf{w}}\left(\sum_{n=1}^{N}\log\sigma(y_n f_n)\right) - \frac{d}{d\mathbf{w}}\left(\frac{1}{2}\mathbf{w}^T\boldsymbol{\Sigma}_w^{-1}\mathbf{w}\right)$$

$$\sum_{n=1}^{N} \frac{d}{d\mathbf{w}} \left(\log \frac{1}{1 + e^{-y_n \mathbf{x}_n^T \mathbf{w}}} \right) - \mathbf{\Sigma}_w^{-1} \mathbf{w}$$

$$\sum_{n=1}^{N} \frac{y_n \mathbf{x}_n}{1 + e^{y_n \mathbf{x}_n^T \mathbf{w}}} - \boldsymbol{\Sigma}_w^{-1} \mathbf{w} = \mathbf{0}$$

y_*

$$p(y_* = 1 | \mathbf{x}_*, \mathbf{w}) = \sigma(\mathbf{x}_*^T \mathbf{w}) = \sigma(f(\mathbf{x}_*)) = \frac{1}{1 + \exp(\mathbf{x}_*^T \mathbf{w})}$$

$$(C_1, \cdots, C_K)$$

$\mathbf{w}_i$

$(i = 1, \cdots, K)$

$$\mathbf{x}^T \mathbf{w}_i$$

$$-\infty < \mathbf{x}^T \mathbf{w}_i < \infty \, (i = 1, \cdots, K)$$

$$p(y \in C_i|\mathbf{x}, \mathbf{W}) = \frac{\exp(\mathbf{x}^T \mathbf{w}_i)}{\sum_{k=1}^{K} \exp(\mathbf{x}^T \mathbf{w}_k)} = \begin{cases} 0 & \mathbf{x}^T \mathbf{w}_i = -\infty \\ 1 & \mathbf{x}^T \mathbf{w}_i = \infty \end{cases}$$

$$\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_K]$$