Tutorial 1: Bayesian Decision Theory

Rui Zhao

rzhao@ee.cuhk.edu.hk

Jan. 16, 2014

Outline

1. Decision to minimize overal expected loss
2. Multivariate Normal Distribution
3. Decision surface for linear machines
   - case 1: $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$
   - case 2: $\mathbf{\Sigma_i} = \mathbf{\Sigma}$

Outline

1. Decision to minimize overal expected loss
2. Multivariate Normal Distribution
3. Decision surface for linear machines
   - case 1: $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$
   - case 2: $\mathbf{\Sigma_i} = \mathbf{\Sigma}$

Outline

1. Decision to minimize overal expected loss
2. Multivariate Normal Distribution
3. Decision surface for linear machines
   - case 1: $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$
   - case 2: $\mathbf{\Sigma_i} = \mathbf{\Sigma}$

Discriminant functions

1. Decision to minimize overal expected loss

Discriminant functions

Decide w_1 if $R(w_1|\mathbf{x}) < R(w_2|\mathbf{x})$, i.e.,

$$\lambda_{11}P(w_1|\mathbf{x}) + \lambda_{12}P(w_2|\mathbf{x}) < \lambda_{21}P(w_1|\mathbf{x}) + \lambda_{22}P(w_2|\mathbf{x})$$

1. Decision to minimize overal expected loss

Discriminant functions

Decide w_1 if $R(w_1|\mathbf{x}) < R(w_2|\mathbf{x})$, i.e.,

$$\lambda_{11}P(w_1|\mathbf{x}) + \lambda_{12}P(w_2|\mathbf{x}) < \lambda_{21}P(w_1|\mathbf{x}) + \lambda_{22}P(w_2|\mathbf{x})$$

$$\Downarrow ?$$

$$\frac{P(\mathbf{x}|w_1)}{P(\mathbf{x}|w_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(w_2)}{P(w_1)}$$

1. Decision to minimize overal expected loss

Expected loss of taking decision w_i

$$R(w_i|\mathbf{x}) = \sum_{j=1}^{2} \lambda_{ij} P(w_j|\mathbf{x}) \quad (1)$$

where λ_{ij} is the loss for deciding w_i when the true class is w_j.

1. Decision to minimize overal expected loss

Expected loss of taking decision w_i

$$R(w_i|\mathbf{x}) = \sum_{j=1}^{2} \lambda_{ij} P(w_j|\mathbf{x}) \tag{1}$$

where λ_{ij} is the loss for deciding w_i when the true class is w_j.

Derive discriminant functions to minimize overal risk

Decide w_1 if $R(w_1|\mathbf{x}) < R(w_2|\mathbf{x})$, i.e.

$$\sum_{j=1}^{2} \lambda_{1j} P(w_j|\mathbf{x}) < \sum_{j=1}^{2} \lambda_{2j} P(w_j|\mathbf{x}) \tag{2}$$

1. Decision to minimize overal expected loss

Bayesian theory

$$P(w_j|\mathbf{x}) \sim P(\mathbf{x}|w_j)P(w_j) \tag{3}$$

1. Decision to minimize overal expected loss

Bayesian theory

$$P(w_j|\mathbf{x}) \sim P(\mathbf{x}|w_j)P(w_j) \tag{3}$$

Decide w_1 if $R(w_1|\mathbf{x}) < R(w_2|\mathbf{x})$, i.e. discriminant function is derived as follows:

$$\sum_{j=1}^{2} \lambda_{1j} P(w_j|\mathbf{x}) < \sum_{j=1}^{2} \lambda_{2j} P(w_j|\mathbf{x}) \tag{4}$$

$$\Rightarrow \lambda_{11}P(w_1|\mathbf{x}) + \lambda_{12}P(w_2|\mathbf{x}) < \lambda_{21}P(w_1|\mathbf{x}) + \lambda_{22}P(w_2|\mathbf{x}) \tag{5}$$

$$\Rightarrow \lambda_{11}P(\mathbf{x}|w_1)P(w_1) + \lambda_{12}P(\mathbf{x}|w_2)P(w_2) \tag{6}$$

$$< \lambda_{21}P(\mathbf{x}|w_1)P(w_1) + \lambda_{22}P(\mathbf{x}|w_2)P(w_2) \tag{7}$$

$$\Rightarrow (\lambda_{12} - \lambda_{22})P(\mathbf{x}|w_2)P(w_2) < (\lambda_{21} - \lambda_{11})P(\mathbf{x}|w_1)P(w_1) \tag{8}$$

$$\Rightarrow \frac{P(\mathbf{x}|w_1)}{P(\mathbf{x}|w_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(w_2)}{P(w_1)}.\# \tag{9}$$

2. Multivariate Normal Distribution

Proposition: the distribution of each variable of a multivariate normal distribution is also a Gaussian.

$$\mathbf{x} \sim \mathcal{N}(\mu, \mathbf{\Sigma})$$

2. Multivariate Normal Distribution

Proposition: the distribution of each variable of a multivariate normal distribution is also a Gaussian.

$$\mathbf{x} \sim \mathcal{N}(\mu, \mathbf{\Sigma})$$

$$\Downarrow ?$$

$$x_i \sim \mathcal{N}(\mu_i, \sigma_{ii}^2)$$

$$p(x_i) = \int \cdots \int p(\mathbf{x}) \cdots dx_{i-1} dx_{i+1} \cdots$$

2. Multivariate Normal Distribution

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

$$p(\mathbf{x}) \sim \mathcal{N}(\mu, \mathbf{\Sigma})$$

- $\mathbf{x} = (x_1, \ldots, x_d)^t$ is the multivariate variable
- $\mu = (\mu_1, \ldots, \mu_d)^t$ is the mean vector
- $\mathbf{\Sigma} = [\sigma_{ij}]$ is the $d \times d$ covariance matrix

2. Multivariate Normal Distribution

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

$$p(\mathbf{x}) \sim \mathcal{N}(\mu, \mathbf{\Sigma})$$

- $\mathbf{x} = (x_1, \ldots, x_d)^t$ is the multivariate variable
- $\mu = (\mu_1, \ldots, \mu_d)^t$ is the mean vector
- $\Sigma = [\sigma_{ij}]$ is the $d \times d$ covariance matrix

2. Multivariate Normal Distribution

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

$$p(\mathbf{x}) \sim \mathcal{N}(\mu, \mathbf{\Sigma})$$

- $\mathbf{x} = (x_1, \ldots, x_d)^t$ is the multivariate variable
- $\mu = (\mu_1, \ldots, \mu_d)^t$ is the mean vector
- $\Sigma = [\sigma_{ij}]$ is the $d \times d$ covariance matrix

2. Multivariate Normal Distribution

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

$$p(\mathbf{x}) \sim \mathcal{N}(\mu, \mathbf{\Sigma})$$

- $\mathbf{x} = (x_1, \ldots, x_d)^t$ is the multivariate variable
- $\mu = (\mu_1, \ldots, \mu_d)^t$ is the mean vector
- $\mathbf{\Sigma} = [\sigma_{ij}]$ is the $d \times d$ covariance matrix

2. Multivariate Normal Distribution

Covariance matrix $\boldsymbol{\Sigma}$

- σ_{ij} measures the covariance between variable x_i and x_j.
- If x_i and x_j are statistically independent, then $\sigma_{ij} = 0$.
- If all variables are independent, then $p(\mathbf{x}) = p(x_1) \cdots p(x_d)$

Covariance matrix $\mathbf{\Sigma}$

- σ_{ij} measures the covariance between variable x_i and x_j.
- If x_i and x_j are statistically independent, then $\sigma_{ij} = 0$.
- If all variables are independent, then $p(\mathbf{x}) = p(x_1) \cdots p(x_d)$

Marginal distribution

The distribution of each variable x_i is also a Gaussian, i.e.
$x_i \sim \mathcal{N}(\mu_i, \sigma_{ii}^2)$
Proof:
Firstly, we consider the linear transform $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$.
It is easy to prove that $\mathbf{x} \sim \mathcal{N}(\mu, \mathbf{\Sigma}) \rightarrow \mathbf{y} \sim \mathcal{N}(\mathbf{A}\mu + \mathbf{b}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^{\mathbf{t}})$.
Let's choose $\mathbf{A} = e_i^t, \mathbf{b} = 0$, the unit vector with i-th entry is 1,
then we have $\mathbf{y} = x_i \sim \mathcal{N}(\mu_i, \sigma_{ii})$. #

3. Decision surface for linear machines

Discriminant function of multivariate normal distribution

$$g_i(x) = \ln p(\mathbf{x}|w_i) + \ln P(w_i) \tag{10}$$

$$p(\mathbf{x}|w_i) \sim \mathcal{N}(\mu_i, \mathbf{\Sigma}_i) \tag{11}$$

Discriminant function of multivariate normal distribution

$$g_i(x) = \ln p(\mathbf{x}|w_i) + \ln P(w_i) \tag{10}$$

$$p(\mathbf{x}|w_i) \sim \mathcal{N}(\mu_i, \mathbf{\Sigma}_i) \tag{11}$$

$$\Rightarrow g_i(x) = -\frac{1}{2}(\mathbf{x} - \mu_i)^t \mathbf{\Sigma}_i^{-1} (\mathbf{x} - \mu_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{\Sigma}_i| + \ln P(w_i)$$

Discriminant function of multivariate normal distribution

$$g_i(x) = \ln p(\mathbf{x}|w_i) + \ln P(w_i) \tag{10}$$

$$p(\mathbf{x}|w_i) \sim \mathcal{N}(\mu_i, \mathbf{\Sigma}_i) \tag{11}$$

$$\Rightarrow g_i(x) = -\frac{1}{2}(\mathbf{x} - \mu_i)^t \mathbf{\Sigma}_i^{-1}(\mathbf{x} - \mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{\Sigma}_i| + \ln P(w_i)$$

For cases that $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$ and $\mathbf{\Sigma_i} = \mathbf{\Sigma}$, the discriminant function is linear (i.e. the classifier is a linear machine), and the decision surfaces are hyperplanes defined by $g_i(\mathbf{x}) = g_j(\mathbf{x})$.

3. Decision surface for linear machines

Decision surface of case $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$

3. Decision surface for linear machines

Decision surface of case $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$

This implies that features are independent with the same variance.

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x} - \mu_i\|^2}{2\sigma^2} + \ln P(w_i) \tag{12}$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} \tag{13}$$

$$\mathbf{w}_i = \frac{\mu_i}{\sigma^2}, \quad \mathbf{w}_{i0} = -\frac{1}{2\sigma^2}\mu_i^t\mu_i + \ln P(w_i) \tag{14}$$

3. Decision surface for linear machines

Decision surface of case $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$

This implies that features are independent with the same variance.

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x} - \mu_i\|^2}{2\sigma^2} + \ln P(w_i) \tag{12}$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} \tag{13}$$

$$\mathbf{w}_i = \frac{\mu_i}{\sigma^2}, \quad \mathbf{w}_{i0} = -\frac{1}{2\sigma^2}\mu_i^t\mu_i + \ln P(w_i) \tag{14}$$

Then the hyperplanes are defined by $g_i(\mathbf{x}) = g_j(\mathbf{x})$, which can be written as

$$\mathbf{w}^t(\mathbf{x} - \mathbf{x}_0) = 0 \tag{15}$$

Decision surface of case $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$

This implies that features are independent with the same variance.

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x} - \mu_i\|^2}{2\sigma^2} + \ln P(w_i) \tag{12}$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} \tag{13}$$

$$\mathbf{w}_i = \frac{\mu_i}{\sigma^2}, \quad \mathbf{w}_{i0} = -\frac{1}{2\sigma^2}\mu_i^t\mu_i + \ln P(w_i) \tag{14}$$

Then the hyperplanes are defined by $g_i(\mathbf{x}) = g_j(\mathbf{x})$, which can be written as

$$\mathbf{w}^t(\mathbf{x} - \mathbf{x}_0) = 0 \tag{15}$$

How to derive $\mathbf{w}$ and $\mathbf{x}_0$?

3. Decision surface for linear machines

Decision surface of case $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{16}$$

3. Decision surface for linear machines

Decision surface of case $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{16}$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \tag{17}$$

Decision surface of case $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{16}$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \tag{17}$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t\mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \tag{18}$$

Decision surface of case $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{16}$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \tag{17}$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t\mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \tag{18}$$

$$\Rightarrow (\frac{\mu_i}{\sigma^2} - \frac{\mu_j}{\sigma^2})^t\mathbf{x} - \Big(\frac{1}{2\sigma^2}\mu_i^t\mu_i - \ln P(w_i) - \frac{1}{2\sigma^2}\mu_j^t\mu_j + \ln P(w_j)\Big) = 0$$

Decision surface of case $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (16)$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \quad (17)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t\mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (18)$$

$$\Rightarrow (\frac{\mu_i}{\sigma^2} - \frac{\mu_j}{\sigma^2})^t\mathbf{x} - \left(\frac{1}{2\sigma^2}\mu_i^t\mu_i - \ln P(w_i) - \frac{1}{2\sigma^2}\mu_j^t\mu_j + \ln P(w_j)\right) = 0$$

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

Decision surface of case $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{16}$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \tag{17}$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t\mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \tag{18}$$

$$\Rightarrow (\frac{\mu_i}{\sigma^2} - \frac{\mu_j}{\sigma^2})^t\mathbf{x} - \Big(\frac{1}{2\sigma^2}\mu_i^t\mu_i - \ln P(w_i) - \frac{1}{2\sigma^2}\mu_j^t\mu_j + \ln P(w_j)\Big) = 0$$

$$\Rightarrow (\mu_i - \mu_j)^t\mathbf{x} - \Big(\frac{1}{2}(\mu_i - \mu_j)^t(\mu_i + \mu_j) - \sigma^2[\ln P(w_i) - \ln P(w_j)]\Big) = 0$$

$$\Rightarrow (\mu_i - \mu_j)^t(\mathbf{x} - \mathbf{x}_0) = 0 \tag{19}$$

Decision surface of case $\mathbf{\Sigma_i} = \sigma^2\mathbf{I}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{16}$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \tag{17}$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t\mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \tag{18}$$

$$\Rightarrow (\frac{\mu_i}{\sigma^2} - \frac{\mu_j}{\sigma^2})^t\mathbf{x} - \left(\frac{1}{2\sigma^2}\mu_i^t\mu_i - \ln P(w_i) - \frac{1}{2\sigma^2}\mu_j^t\mu_j + \ln P(w_j)\right) = 0$$

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

$$\Rightarrow (\mu_i - \mu_j)^t(\mathbf{x} - \mathbf{x}_0) = 0 \tag{19}$$

$$\Rightarrow \mathbf{w} = \mu_i - \mu_j, \quad \mathbf{x}_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{\|\mu_i - \mu_j\|^2} \ln \frac{P(w_i)}{P(w_j)}(\mu_i - \mu_j).\#$$

Decision surface of case $\mathbf{\Sigma_i = \Sigma}$

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)^t \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu_i) + \ln P(w_i) \tag{20}$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} \tag{21}$$

$$\mathbf{w}_i = \mathbf{\Sigma}^{-1}\mu_i, \quad \mathbf{w}_{i0} = -\frac{1}{2}\mu_i^t \mathbf{\Sigma}^{-1}\mu_i + \ln P(w_i) \tag{22}$$

Decision surface of case $\mathbf{\Sigma_i = \Sigma}$

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)^t \mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_i) + \ln P(w_i) \tag{20}$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} \tag{21}$$

$$\mathbf{w}_i = \mathbf{\Sigma}^{-1}\mu_i, \quad \mathbf{w}_{i0} = -\frac{1}{2}\mu_i^t \mathbf{\Sigma}^{-1}\mu_i + \ln P(w_i) \tag{22}$$

Then the hyperplanes are defined by $g_i(\mathbf{x}) = g_j(\mathbf{x})$, which can be written as

$$\mathbf{w}^t(\mathbf{x} - \mathbf{x}_0) = 0 \tag{23}$$

Decision surface of case $\mathbf{\Sigma_i = \Sigma}$

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)^t \mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_i) + \ln P(w_i) \tag{20}$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} \tag{21}$$

$$\mathbf{w}_i = \mathbf{\Sigma}^{-1}\mu_i, \quad \mathbf{w}_{i0} = -\frac{1}{2}\mu_i^t \mathbf{\Sigma}^{-1}\mu_i + \ln P(w_i) \tag{22}$$

Then the hyperplanes are defined by $g_i(\mathbf{x}) = g_j(\mathbf{x})$, which can be written as

$$\mathbf{w}^t(\mathbf{x} - \mathbf{x}_0) = 0 \tag{23}$$

How to derive $\mathbf{w}$ and $\mathbf{x}_0$?

3. Decision surface for linear machines

Decision surface of case $\mathbf{\Sigma_i} = \mathbf{\Sigma}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{24}$$

Decision surface of case $\mathbf{\Sigma_i} = \mathbf{\Sigma}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{24}$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \tag{25}$$

3. Decision surface for linear machines

Decision surface of case $\mathbf{\Sigma_i} = \mathbf{\Sigma}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{24}$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \tag{25}$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t\mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \tag{26}$$

3. Decision surface for linear machines

Decision surface of case $\mathbf{\Sigma_i} = \mathbf{\Sigma}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{24}$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \tag{25}$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t\mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \tag{26}$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}\mu_i - \mathbf{\Sigma}^{-1}\mu_j)^t\mathbf{x} \tag{27}$$

3. Decision surface for linear machines

Decision surface of case $\mathbf{\Sigma_i} = \mathbf{\Sigma}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{24}$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \tag{25}$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t\mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \tag{26}$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}\mu_i - \mathbf{\Sigma}^{-1}\mu_j)^t\mathbf{x} - \Big(\frac{1}{2}\mu_i^t\mathbf{\Sigma}^{-1}\mu_i - \ln P(w_i) - \frac{1}{2}\mu_j^t\mathbf{\Sigma}^{-1}\mu_j + \ln P(w_j)\Big) = 0 \tag{27}$$

Decision surface of case $\mathbf{\Sigma_i} = \mathbf{\Sigma}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{24}$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \tag{25}$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t\mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \tag{26}$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}\mu_i - \mathbf{\Sigma}^{-1}\mu_j)^t\mathbf{x} \tag{27}$$

$$- \left(\frac{1}{2}\mu_i^t\mathbf{\Sigma}^{-1}\mu_i - \ln P(w_i) - \frac{1}{2}\mu_j^t\mathbf{\Sigma}^{-1}\mu_j + \ln P(w_j)\right) = 0$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}(\mu_i - \mu_j))^t\mathbf{x} - \left(\frac{1}{2}\mu_i^t\mathbf{\Sigma}^{-1}\mu_i - \frac{1}{2}\mu_j^t\mathbf{\Sigma}^{-1}\mu_j - \ln\frac{P(w_i)}{P(w_j)}\right) = 0$$

3. Decision surface for linear machines

Decision surface of case $\mathbf{\Sigma_i} = \mathbf{\Sigma}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{24}$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \tag{25}$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \tag{26}$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}\mu_i - \mathbf{\Sigma}^{-1}\mu_j)^t \mathbf{x} \tag{27}$$

$$- \left(\frac{1}{2}\mu_i^t \mathbf{\Sigma}^{-1}\mu_i - \ln P(w_i) - \frac{1}{2}\mu_j^t \mathbf{\Sigma}^{-1}\mu_j + \ln P(w_j) \right) = 0$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}(\mu_i - \mu_j))^t \mathbf{x} - \left(\frac{1}{2}\mu_i^t \mathbf{\Sigma}^{-1}\mu_i - \frac{1}{2}\mu_j^t \mathbf{\Sigma}^{-1}\mu_j - \ln \frac{P(w_i)}{P(w_j)} \right) = 0$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}(\mu_i - \mu_j))^t \mathbf{x} - \left(\frac{1}{2}(\mu_i - \mu_j)^t \mathbf{\Sigma}^{-1}(\mu_i + \mu_j) - \ln \frac{P(w_i)}{P(w_j)} \right) = 0$$

3. Decision surface for linear machines

Decision surface of case $\mathbf{\Sigma_i} = \mathbf{\Sigma}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{24}$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \tag{25}$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t\mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \tag{26}$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}\mu_i - \mathbf{\Sigma}^{-1}\mu_j)^t\mathbf{x} \tag{27}$$

$$- \Big(\frac{1}{2}\mu_i^t\mathbf{\Sigma}^{-1}\mu_i - \ln P(w_i) - \frac{1}{2}\mu_j^t\mathbf{\Sigma}^{-1}\mu_j + \ln P(w_j)\Big) = 0$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}(\mu_i - \mu_j))^t\mathbf{x} - \Big(\frac{1}{2}\mu_i^t\mathbf{\Sigma}^{-1}\mu_i - \frac{1}{2}\mu_j^t\mathbf{\Sigma}^{-1}\mu_j - \ln\frac{P(w_i)}{P(w_j)}\Big) = 0$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}(\mu_i - \mu_j))^t\mathbf{x} - \Big(\frac{1}{2}(\mu_i - \mu_j)^t\mathbf{\Sigma}^{-1}(\mu_i + \mu_j) - \ln\frac{P(w_i)}{P(w_j)}\Big) = 0$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}(\mu_i - \mu_j))^t(\mathbf{x} - \mathbf{x}_0) = 0 \tag{28}$$

Decision surface of case $\mathbf{\Sigma_i} = \mathbf{\Sigma}$: deriving $\mathbf{w}$ and $\mathbf{x}_0$ ...

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \tag{24}$$

$$\Rightarrow \mathbf{w}_i^t\mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t\mathbf{x} + \mathbf{w}_{j0} \tag{25}$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t\mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \tag{26}$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}\mu_i - \mathbf{\Sigma}^{-1}\mu_j)^t\mathbf{x} \tag{27}$$

$$- \left(\frac{1}{2}\mu_i^t\mathbf{\Sigma}^{-1}\mu_i - \ln P(w_i) - \frac{1}{2}\mu_j^t\mathbf{\Sigma}^{-1}\mu_j + \ln P(w_j)\right) = 0$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}(\mu_i - \mu_j))^t\mathbf{x} - \left(\frac{1}{2}\mu_i^t\mathbf{\Sigma}^{-1}\mu_i - \frac{1}{2}\mu_j^t\mathbf{\Sigma}^{-1}\mu_j - \ln\frac{P(w_i)}{P(w_j)}\right) = 0$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}(\mu_i - \mu_j))^t\mathbf{x} - \left(\frac{1}{2}(\mu_i - \mu_j)^t\mathbf{\Sigma}^{-1}(\mu_i + \mu_j) - \ln\frac{P(w_i)}{P(w_j)}\right) = 0$$

$$\Rightarrow (\mathbf{\Sigma}^{-1}(\mu_i - \mu_j))^t(\mathbf{x} - \mathbf{x}_0) = 0 \tag{28}$$

$$\Rightarrow \mathbf{w} = \mathbf{\Sigma}^{-1}(\mu_i - \mu_j), \quad \mathbf{x}_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\ln[P(w_i)/P(w_j)](\mu_i - \mu_j)}{(\mu_i - \mu_j)^t\mathbf{\Sigma}^{-1}(\mu_i - \mu_j)}.\#$$