

Tutorial 1: Bayesian Decision Theory

Rui Zhao

rzhao@ee.cuhk.edu.hk

Jan. 16, 2014

- 1 ML estimate of multivariate Gaussian
- 2 Multivariate Normal Distribution
- 3 Decision surface for linear machines
 - case 1: $\Sigma_i = \sigma^2 \mathbf{I}$
 - case 2: $\Sigma_i = \Sigma$

- 1 ML estimate of multivariate Gaussian
- 2 Multivariate Normal Distribution
- 3 Decision surface for linear machines
 - case 1: $\Sigma_i = \sigma^2 \mathbf{I}$
 - case 2: $\Sigma_i = \Sigma$

- 1 ML estimate of multivariate Gaussian
- 2 Multivariate Normal Distribution
- 3 Decision surface for linear machines
 - case 1: $\Sigma_i = \sigma^2 \mathbf{I}$
 - case 2: $\Sigma_i = \Sigma$

1. ML estimate of multivariate Gaussian

Gaussian Case: unknown μ and Σ

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k \quad (1)$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^n (\mathbf{x}_k - \hat{\mu})(\mathbf{x}_k - \hat{\mu})^t \quad (2)$$

1. Decision to minimize overall expected loss

Expected loss of taking decision w_i

$$R(w_i|\mathbf{x}) = \sum_{j=1}^2 \lambda_{ij} P(w_j|\mathbf{x}) \quad (3)$$

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

1. Decision to minimize overall expected loss

Expected loss of taking decision w_i

$$R(w_i|\mathbf{x}) = \sum_{j=1}^2 \lambda_{ij} P(w_j|\mathbf{x}) \quad (3)$$

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

Derive discriminant functions to minimize overall risk

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

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

1. Decision to minimize overall expected loss

Bayesian theory

$$P(w_j|\mathbf{x}) \sim P(\mathbf{x}|w_j)P(w_j) \quad (5)$$

1. Decision to minimize overall expected loss

Bayesian theory

$$P(w_j|\mathbf{x}) \sim P(\mathbf{x}|w_j)P(w_j) \quad (5)$$

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

$$\sum_{j=1}^2 \lambda_{1j} P(w_j|\mathbf{x}) < \sum_{j=1}^2 \lambda_{2j} P(w_j|\mathbf{x}) \quad (6)$$

$$\Rightarrow \lambda_{11}P(w_1|\mathbf{x}) + \lambda_{12}P(w_2|\mathbf{x}) < \lambda_{21}P(w_1|\mathbf{x}) + \lambda_{22}P(w_2|\mathbf{x}) \quad (7)$$

$$\Rightarrow \lambda_{11}P(\mathbf{x}|w_1)P(w_1) + \lambda_{12}P(\mathbf{x}|w_2)P(w_2) \quad (8)$$

$$< \lambda_{21}P(\mathbf{x}|w_1)P(w_1) + \lambda_{22}P(\mathbf{x}|w_2)P(w_2) \quad (9)$$

$$\Rightarrow (\lambda_{12} - \lambda_{22})P(\mathbf{x}|w_2)P(w_2) < (\lambda_{21} - \lambda_{11})P(\mathbf{x}|w_1)P(w_1) \quad (10)$$

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

2. Multivariate Normal Distribution

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

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

2. Multivariate Normal Distribution

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

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



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

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

2. Multivariate Normal Distribution

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

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

2. Multivariate Normal Distribution

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

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

2. Multivariate Normal Distribution

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

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

2. Multivariate Normal Distribution

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

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

2. Multivariate Normal Distribution

Covariance matrix Σ

- 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 $p(\mathbf{x}) = p(x_1) \cdots p(x_d)$

2. Multivariate Normal Distribution

Covariance matrix Σ

- 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 $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma) \rightarrow \mathbf{y} \sim \mathcal{N}(\mathbf{A}\mu + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}^t)$.

Let's choose $\mathbf{A} = \mathbf{e}_i^t$, $\mathbf{b} = \mathbf{0}$, 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

$$g_i(x) = \ln p(\mathbf{x}|w_i) + \ln P(w_i) \quad (12)$$

$$p(\mathbf{x}|w_i) \sim \mathcal{N}(\mu_i, \Sigma_i) \quad (13)$$

3. Decision surface for linear machines

Discriminant function of multivariate normal distribution

$$g_i(x) = \ln p(\mathbf{x}|w_i) + \ln P(w_i) \quad (12)$$

$$p(\mathbf{x}|w_i) \sim \mathcal{N}(\mu_i, \Sigma_i) \quad (13)$$

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

3. Decision surface for linear machines

Discriminant function of multivariate normal distribution

$$g_i(x) = \ln p(\mathbf{x}|w_i) + \ln P(w_i) \quad (12)$$

$$p(\mathbf{x}|w_i) \sim \mathcal{N}(\mu_i, \Sigma_i) \quad (13)$$

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

For cases that $\Sigma_i = \sigma^2 \mathbf{I}$ and $\Sigma_i = \Sigma$, 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 $\Sigma_i = \sigma^2 \mathbf{I}$

3. Decision surface for linear machines

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

This implies that features are independent with the same variance.

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x} - \mu_i\|^2}{2\sigma^2} + \ln P(w_i) \quad (14)$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} \quad (15)$$

$$\mathbf{w}_i = \frac{\mu_i}{\sigma^2}, \quad \mathbf{w}_{i0} = -\frac{1}{2\sigma^2} \mu_i^t \mu_i + \ln P(w_i) \quad (16)$$

3. Decision surface for linear machines

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

This implies that features are independent with the same variance.

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x} - \mu_i\|^2}{2\sigma^2} + \ln P(w_i) \quad (14)$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} \quad (15)$$

$$\mathbf{w}_i = \frac{\mu_i}{\sigma^2}, \quad \mathbf{w}_{i0} = -\frac{1}{2\sigma^2} \mu_i^t \mu_i + \ln P(w_i) \quad (16)$$

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

$$\mathbf{w}^t (\mathbf{x} - \mathbf{x}_0) = 0 \quad (17)$$

3. Decision surface for linear machines

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

This implies that features are independent with the same variance.

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x} - \mu_i\|^2}{2\sigma^2} + \ln P(w_i) \quad (14)$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} \quad (15)$$

$$\mathbf{w}_i = \frac{\mu_i}{\sigma^2}, \quad \mathbf{w}_{i0} = -\frac{1}{2\sigma^2} \mu_i^t \mu_i + \ln P(w_i) \quad (16)$$

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

$$\mathbf{w}^t (\mathbf{x} - \mathbf{x}_0) = 0 \quad (17)$$

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

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (18)$$

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (18)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (19)$$

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (18)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (19)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (20)$$

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (18)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (19)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (20)$$

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

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (18)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (19)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (20)$$

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

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

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (18)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (19)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (20)$$

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

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

$$\Rightarrow (\mu_i - \mu_j)^t (\mathbf{x} - \mathbf{x}_0) = 0 \quad (21)$$

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (18)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (19)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (20)$$

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

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

$$\Rightarrow (\mu_i - \mu_j)^t (\mathbf{x} - \mathbf{x}_0) = 0 \quad (21)$$

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

3. Decision surface for linear machines

Decision surface of case $\Sigma_i = \Sigma$

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)^t \Sigma^{-1}(\mathbf{x} - \mu_i) + \ln P(w_i) \quad (22)$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} \quad (23)$$

$$\mathbf{w}_i = \Sigma^{-1} \mu_i, \quad \mathbf{w}_{i0} = -\frac{1}{2} \mu_i^t \Sigma^{-1} \mu_i + \ln P(w_i) \quad (24)$$

3. Decision surface for linear machines

Decision surface of case $\Sigma_i = \Sigma$

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)^t \Sigma^{-1}(\mathbf{x} - \mu_i) + \ln P(w_i) \quad (22)$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} \quad (23)$$

$$\mathbf{w}_i = \Sigma^{-1} \mu_i, \quad \mathbf{w}_{i0} = -\frac{1}{2} \mu_i^t \Sigma^{-1} \mu_i + \ln P(w_i) \quad (24)$$

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

$$\mathbf{w}^t(\mathbf{x} - \mathbf{x}_0) = 0 \quad (25)$$

3. Decision surface for linear machines

Decision surface of case $\Sigma_i = \Sigma$

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)^t \Sigma^{-1}(\mathbf{x} - \mu_i) + \ln P(w_i) \quad (22)$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} \quad (23)$$

$$\mathbf{w}_i = \Sigma^{-1} \mu_i, \quad \mathbf{w}_{i0} = -\frac{1}{2} \mu_i^t \Sigma^{-1} \mu_i + \ln P(w_i) \quad (24)$$

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

$$\mathbf{w}^t(\mathbf{x} - \mathbf{x}_0) = 0 \quad (25)$$

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

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (26)$$

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (26)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (27)$$

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (26)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (27)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (28)$$

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (26)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (27)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (28)$$

$$\Rightarrow (\Sigma^{-1} \mu_i - \Sigma^{-1} \mu_j)^t \mathbf{x} \quad (29)$$

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (26)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (27)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (28)$$

$$\Rightarrow (\Sigma^{-1} \mu_i - \Sigma^{-1} \mu_j)^t \mathbf{x} \quad (29)$$

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

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (26)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (27)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (28)$$

$$\Rightarrow (\Sigma^{-1} \mu_i - \Sigma^{-1} \mu_j)^t \mathbf{x} \quad (29)$$

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

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

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (26)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (27)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (28)$$

$$\Rightarrow (\Sigma^{-1} \mu_i - \Sigma^{-1} \mu_j)^t \mathbf{x} \quad (29)$$

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

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

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

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (26)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (27)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (28)$$

$$\Rightarrow (\Sigma^{-1} \mu_i - \Sigma^{-1} \mu_j)^t \mathbf{x} \quad (29)$$

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

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

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

$$\Rightarrow (\Sigma^{-1} (\mu_i - \mu_j))^t (\mathbf{x} - \mathbf{x}_0) = 0 \quad (30)$$

3. Decision surface for linear machines

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

$$g_i(\mathbf{x}) = g_j(\mathbf{x}) \quad (26)$$

$$\Rightarrow \mathbf{w}_i^t \mathbf{x} + \mathbf{w}_{i0} = \mathbf{w}_j^t \mathbf{x} + \mathbf{w}_{j0} \quad (27)$$

$$\Rightarrow (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (\mathbf{w}_{i0} - \mathbf{w}_{j0}) = 0 \quad (28)$$

$$\Rightarrow (\Sigma^{-1} \mu_i - \Sigma^{-1} \mu_j)^t \mathbf{x} \quad (29)$$

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

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

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

$$\Rightarrow (\Sigma^{-1} (\mu_i - \mu_j))^t (\mathbf{x} - \mathbf{x}_0) = 0 \quad (30)$$

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