# Big Data Analysis
## (MA60306)

Bibhas Adhikari

Spring 2022-23, IIT Kharagpur

Lecture 16
March 9, 2023

# Review

LDA 2-class classification

$\rightarrow$ Classes: $\Pi_1, \Pi_2$ and prior probabilities - $P(\mathbf{X} \in \Pi_i) = \pi_i, \ i = 1, 2$

# Review

LDA 2-class classification

- $\rightarrow$ Classes: $\Pi_1, \Pi_2$ and prior probabilities - $P(\mathbf{X} \in \Pi_i) = \pi_i, \; i = 1, 2$
- $\rightarrow$ Conditional probabilities - $P(\mathbf{X} = \mathbf{x} \,|\, \mathbf{X} \in \Pi_i) = f_i(\mathbf{x}), i = 1, 2$

# Review

LDA 2-class classification

- $\rightarrow$ Classes: $\Pi_1, \Pi_2$ and prior probabilities - $P(\mathbf{X} \in \Pi_i) = \pi_i,\ i = 1, 2$
- $\rightarrow$ Conditional probabilities - $P(\mathbf{X} = \mathbf{x} \mid \mathbf{X} \in \Pi_i) = f_i(\mathbf{x}), i = 1, 2$
- $\rightarrow$ posterior probabilities -

$$p(\Pi_i|\mathbf{x}) = \frac{f_i(\mathbf{x})\pi_i}{f_1(\mathbf{x})\pi_1 + f_2(\mathbf{x})\pi_2}$$

# Review

LDA 2-class classification

- $\rightarrow$ Classes: $\Pi_1, \Pi_2$ and prior probabilities - $P(\mathbf{X} \in \Pi_i) = \pi_i,\ i = 1, 2$
- $\rightarrow$ Conditional probabilities - $P(\mathbf{X} = \mathbf{x} \,|\, \mathbf{X} \in \Pi_i) = f_i(\mathbf{x}), i = 1, 2$
- $\rightarrow$ posterior probabilities -

$$p(\Pi_i|\mathbf{x}) = \frac{f_i(\mathbf{x})\pi_i}{f_1(\mathbf{x})\pi_1 + f_2(\mathbf{x})\pi_2}$$

- $\rightarrow$ Bayes's rule classifier: Assign $\mathbf{x}$ to $\Pi_1$ if

$$r = \frac{p(\Pi_1|\mathbf{x})}{p(\Pi_2|\mathbf{x})} > 1 \text{ i.e. } \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} > \frac{\pi_2}{\pi_1}$$

and assign $\mathbf{x}$ to $\Pi_2$ otherwise.

# Review

- → Gaussian LDA: $f_1(x)$ and $f_2(x)$ be multivariate Gaussian having arbitrary mean vectors and a common covariance matrix $\Sigma$: (what is the geometry?)

$$f_1(\cdot) \sim \mathcal{N}_d(\mu_1, \Sigma), \text{ and } f_2(\cdot) \sim \mathcal{N}_d(\mu_2, \Sigma).$$

- → $d$-variate Gaussian (Normal) distribution with mean vector $\boldsymbol{\mu}$ and positive-definite $d \times d$ covariance matrix $\boldsymbol{\Sigma}$ is

$$f(\mathbf{x}) = (2\pi)^{-d/2}|\boldsymbol{\Sigma}|^{-1/2}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

# Review

→ Gaussian LDA: $f_1(x)$ and $f_2(x)$ be multivariate Gaussian having arbitrary mean vectors and a common covariance matrix $\Sigma$: (what is the geometry?)

$$f_1(\cdot) \sim \mathcal{N}_d(\mu_1, \Sigma), \text{ and } f_2(\cdot) \sim \mathcal{N}_d(\mu_2, \Sigma).$$

→ $d$-variate Gaussian (Normal) distribution with mean vector $\boldsymbol{\mu}$ and positive-definite $d \times d$ covariance matrix $\boldsymbol{\Sigma}$ is

$$f(\mathbf{x}) = (2\pi)^{-d/2}|\boldsymbol{\Sigma}|^{-1/2}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

→ Classification rule (Gaussian LDA): Assign $\mathbf{x}$ to $\Pi_1$ if $L(\mathbf{x}) > 0$, otherwise assign $\mathbf{x}$ to $\Pi_2$, where $L(\mathbf{x}) = \log_e\left\{\frac{f_1(\mathbf{x})\pi_1}{f_2(\mathbf{z})\pi_2}\right\} = b_0 + \mathbf{b}^T\mathbf{x}$, with

$$\begin{aligned} \mathbf{b} &= \Sigma^{-1}(\mu_1 - \mu_2) \\ b_0 &= -\frac{1}{2}\left\{\mu_1^T\Sigma^{-1}\mu_1 - \mu_2^T\Sigma^{-1}\mu_2\right\} + \log_e(\pi_2/\pi_1) \end{aligned}$$

# LDA

Squared Mahalanobis distance The squared Mahalanobis distance between $\Pi_1$ and $\Pi_2$ is defined as

$$\triangle^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).$$

# LDA

Squared Mahalanobis distance The squared Mahalanobis distance between $\Pi_1$ and $\Pi_2$ is defined as

$$\triangle^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).$$

Question What does Mahalanobis distance measure?

# LDA

Squared Mahalanobis distance The squared Mahalanobis distance between $\Pi_1$ and $\Pi_2$ is defined as

$$\triangle^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).$$

Question What does Mahalanobis distance measure?

Recall Let $X$ be a random matrix. For the random matrix $Y = AXB^T + C$, where $A, B, C$ are compatible matrices, $\mathbb{E}(Y) = A\mathbb{E}(X)B^T$ and the covariance matrix of $vec(Y)$ is $\Sigma_{YY} = (A \otimes B)\Sigma_{XX}(A \otimes B)^T$

# LDA

Squared Mahalanobis distance The squared Mahalanobis distance between $\Pi_1$ and $\Pi_2$ is defined as

$$\triangle^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).$$

Question What does Mahalanobis distance measure?

Recall Let $X$ be a random matrix. For the random matrix $Y = AXB^T + C$, where $A, B, C$ are compatible matrices, $\mathbb{E}(Y) = A\mathbb{E}(X)B^T$ and the covariance matrix of $vec(Y)$ is $\Sigma_{YY} = (A \otimes B)\Sigma_{XX}(A \otimes B)^T$

Set $U = \mathbf{b}^T\mathbf{x}$, which is a random variable. Then

$$\begin{aligned} \mathbb{E}(U|\mathbf{x} \in \Pi_i) &= \mathbf{b}^T\mu_i = (\mu_1 - \mu_2)^T\Sigma^{-1}\mu_i \\ Var(U|\mathbf{x} \in \Pi_i) &= \mathbf{b}^T\Sigma\mathbf{b} = (\mu_1 - \mu_2)^T\Sigma^{-1}\Sigma\Sigma^{-1}(\mu_1 - \mu_2) = \triangle^2. \end{aligned}$$

# LDA

Squared Mahalanobis distance The squared Mahalanobis distance between $\Pi_1$ and $\Pi_2$ is defined as

$$\triangle^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).$$

Question What does Mahalanobis distance measure?
Recall Let $X$ be a random matrix. For the random matrix $Y = AXB^T + C$, where $A, B, C$ are compatible matrices, $\mathbb{E}(Y) = A\mathbb{E}(X)B^T$ and the covariance matrix of $vec(Y)$ is $\Sigma_{YY} = (A \otimes B)\Sigma_{XX}(A \otimes B)^T$

Set $U = \mathbf{b}^T \mathbf{x}$, which is a random variable. Then

$$\begin{aligned} \mathbb{E}(U|\mathbf{x} \in \Pi_i) &= \mathbf{b}^T \mu_i = (\mu_1 - \mu_2)^T \Sigma^{-1} \mu_i \\ Var(U|\mathbf{x} \in \Pi_i) &= \mathbf{b}^T \Sigma \mathbf{b} = (\mu_1 - \mu_2)^T \Sigma^{-1} \Sigma \Sigma^{-1} (\mu_1 - \mu_2) = \triangle^2. \end{aligned}$$

Let $R_1, R_2$ be the regions given by the classification rule. Then the total misclassification probability:

$$P(\mathbf{x} \in R_2 | \mathbf{x} \in \Pi_1)\pi_1 + P(\mathbf{x} \in R_1 | \mathbf{x} \in \Pi_2)\pi_2$$

# LDA

Now

$$P(\mathbf{x} \in R_2 | \mathbf{x} \in \Pi_1) = P(L(\mathbf{x}) < 0 | \mathbf{x} \in \Pi_1).$$

# LDA

Now

$$P(\mathbf{x} \in R_2 | \mathbf{x} \in \Pi_1) = P(L(\mathbf{x}) < 0 | \mathbf{x} \in \Pi_1).$$

Note that $L(\mathbf{x}) = b_0 + U$. (what is the distribution of $U$?)

# LDA

Now

$$P(\mathbf{x} \in R_2 | \mathbf{x} \in \Pi_1) = P(L(\mathbf{x}) < 0 | \mathbf{x} \in \Pi_1).$$

Note that $L(\mathbf{x}) = b_0 + U$. (what is the distribution of $U$?) Also,

$$Z = \frac{U - \mathbb{E}(U|\mathbf{x} \in \Pi_i)}{\sqrt{var(U|\mathbf{x} \in \Pi_i)}} \sim \mathcal{N}(0, 1)$$

# LDA

Now

$$P(\mathbf{x} \in R_2 | \mathbf{x} \in \Pi_1) = P(L(\mathbf{x}) < 0 | \mathbf{x} \in \Pi_1).$$

Note that $L(\mathbf{x}) = b_0 + U$. (what is the distribution of $U$?) Also,

$$Z = \frac{U - \mathbb{E}(U|\mathbf{x} \in \Pi_i)}{\sqrt{var(U|\mathbf{x} \in \Pi_i)}} \sim \mathcal{N}(0,1)$$

Then from using the expressions for $\mathbb{E}(U|\mathbf{x} \in \Pi_i)$, $Var(U|\mathbf{x} \in \Pi_i)$, and $b_0$ as derived above,

$$\begin{aligned} P(L(\mathbf{x}) < 0 | \mathbf{x} \in \Pi_1) &= P(U < -b_0 | \mathbf{x} \in \Pi_1) \\ &= P(Z < \frac{-b_0 - (\mu_1 - \mu_2)^T \Sigma^{-1} \mu_1}{\triangle}) \\ &= P(Z < -\frac{\triangle}{2} - \frac{1}{\triangle} \log_e \frac{\pi_2}{\pi_1}) \\ &= \Phi\left(-\frac{\triangle}{2} - \frac{1}{\triangle} \log_e \frac{\pi_2}{\pi_1}\right) \end{aligned}$$

# LDA

Similarly we can obtain

$$
\begin{aligned}
P(\mathbf{x} \in R_1 | \mathbf{x} \in \Pi_2) &= P(L(\mathbf{x}) > 0 | \mathbf{x} \in \Pi_2) \\
&= P(Z < -\frac{\triangle}{2} - \frac{1}{\triangle} \log_e \frac{\pi_2}{\pi_1}) \\
&= \Phi\left(-\frac{\triangle}{2} + \frac{1}{\triangle} \log_e \frac{\pi_2}{\pi_1}\right)
\end{aligned}
$$

# LDA

Similarly we can obtain

$$\begin{aligned} P(\mathbf{x} \in R_1 | \mathbf{x} \in \Pi_2) &= P(L(\mathbf{x}) > 0 | \mathbf{x} \in \Pi_2) \\ &= P(Z < -\frac{\triangle}{2} - \frac{1}{\triangle} \log_e \frac{\pi_2}{\pi_1}) \\ &= \Phi\left(-\frac{\triangle}{2} + \frac{1}{\triangle} \log_e \frac{\pi_2}{\pi_1}\right) \end{aligned}$$

If $\pi_1 = \pi_2 = 1/2$ then

$$P(\mathbf{X} \in R_2 | \mathbf{X} \in \Pi_1) = P(\mathbf{X} \in R_1 | \mathbf{X} \in \Pi_2) = \Phi(-\triangle/2)$$

# LDA

Similarly we can obtain

$$
\begin{aligned}
P(\mathbf{x} \in R_1 | \mathbf{x} \in \Pi_2) &= P(L(\mathbf{x}) > 0 | \mathbf{x} \in \Pi_2) \\
&= P(Z < -\frac{\triangle}{2} - \frac{1}{\triangle} \log_e \frac{\pi_2}{\pi_1}) \\
&= \Phi\left(-\frac{\triangle}{2} + \frac{1}{\triangle} \log_e \frac{\pi_2}{\pi_1}\right)
\end{aligned}
$$

If $\pi_1 = \pi_2 = 1/2$ then

$$P(\mathbf{X} \in R_2 | \mathbf{X} \in \Pi_1) = P(\mathbf{X} \in R_1 | \mathbf{X} \in \Pi_2) = \Phi(-\triangle/2)$$

Observation Since miscalculation probability depends on $\triangle$, we can write the probability of miscalculation as $P(\triangle)$. Plotting the graph for $\pi_1 = \pi_2 = 1/2$, what is your conclusion?

# LDA

Question How do we implement the method in real data?

# LDA

Question How do we implement the method in real data?

Observations

→ Note that $\mu_1, \mu_1, \mathbf{\Sigma}$ are not known

# LDA

Question How do we implement the method in real data?

Observations

- $\rightarrow$ Note that $\mu_1, \mu_1, \mathbf{\Sigma}$ are not known
- $\rightarrow$ In general there are $2d + d(d+2)$ distinct parameters in $\mu_1, \mu_2, \mathbf{\Sigma}$ that can possibly be estimated from learning the data

# LDA

Question How do we implement the method in real data?

Observations

- → Note that $\mu_1, \mu_1, \mathbf{\Sigma}$ are not known
- → In general there are $2d + d(d+2)$ distinct parameters in $\mu_1, \mu_2, \mathbf{\Sigma}$ that can possibly be estimated from learning the data
- → Suppose we have a random sample $\mathbf{X}_{1j}, 1 \leq j \leq n_1$, and $\mathbf{X}_{2l}, 1 \leq l \leq n_2$ with values $\mathbf{x}_{1j}$ and $\mathbf{x}_{2l}$ from $\Pi_1$ and $\Pi_2$ respectively

# LDA

Question How do we implement the method in real data?

Observations

- → Note that $\mu_1, \mu_1, \mathbf{\Sigma}$ are not known
- → In general there are $2d + d(d+2)$ distinct parameters in $\mu_1, \mu_2, \mathbf{\Sigma}$ that can possibly be estimated from learning the data
- → Suppose we have a random sample $\mathbf{X}_{1j}, 1 \leq j \leq n_1$, and $\mathbf{X}_{2l}, 1 \leq l \leq n_2$ with values $\mathbf{x}_{1j}$ and $\mathbf{x}_{2l}$ from $\Pi_1$ and $\Pi_2$ respectively

Sampling methods from a population

- → Mixture sampling - a sample of $n = n_1 + n_2$ is selected so that $n_1$ and $n_2$ are randomly selected

# LDA

Question How do we implement the method in real data?

Observations

- → Note that $\mu_1, \mu_1, \mathbf{\Sigma}$ are not known
- → In general there are $2d + d(d+2)$ distinct parameters in $\mu_1, \mu_2, \mathbf{\Sigma}$ that can possibly be estimated from learning the data
- → Suppose we have a random sample $\mathbf{X}_{1j}, 1 \leq j \leq n_1$, and $\mathbf{X}_{2l}, 1 \leq l \leq n_2$ with values $\mathbf{x}_{1j}$ and $\mathbf{x}_{2l}$ from $\Pi_1$ and $\Pi_2$ respectively

Sampling methods from a population

- → Mixture sampling - a sample of $n = n_1 + n_2$ is selected so that $n_1$ and $n_2$ are randomly selected
- → Separate sampling - a sample of $n_i$ is randomly selected from $\Pi_i, i = 1, 2$ and $n = n_1 + n_2$

# LDA

Estimation of parameters The ML estimates of $\mu_i$, $i = 1, 2$ and $\boldsymbol{\Sigma}$ are

# LDA

Estimation of parameters The ML estimates of $\mu_i,\ i = 1, 2$ and $\mathbf{\Sigma}$ are

$$\begin{aligned} \widehat{\mu}_i &= \bar{\mathbf{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{x}_{ij}, i = 1, 2 \text{ and} \\ \widehat{\Sigma} &= \frac{1}{n}S,\ S = S_1 + S_2,\ S_i = \sum_{j=1}^{n_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T \end{aligned}$$

# LDA

Estimation of parameters The ML estimates of $\mu_i,\ i = 1, 2$ and $\mathbf{\Sigma}$ are

$$\begin{aligned}
\widehat{\mu}_i &= \overline{\mathbf{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{x}_{ij}, i = 1, 2 \text{ and} \\
\widehat{\Sigma} &= \frac{1}{n}S,\ S = S_1 + S_2,\ S_i = \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)^T
\end{aligned}$$

Note that for unbiased estimator of $\mathbf{\Sigma}$, we can divide $S$ by its degree of freedom $n - 2 = n_1 + n_2 - 2$ rather than $n$ to make $\widehat{\mathbf{\Sigma}}$

# LDA

The probabilities $\pi_1, \pi_2$ can be chosen based on past experiences or can be estimated as

$$\widehat{\pi}_i = \frac{n_i}{n}, i = 1, 2$$

# LDA

The probabilities $\pi_1, \pi_2$ can be chosen based on past experiences or can be estimated as

$$\widehat{\pi}_i = \frac{n_i}{n}, i = 1, 2$$

Then $\widehat{L}(\mathbf{x}) = \widehat{b}_0 + \widehat{\mathbf{b}}^T\mathbf{x}$, where

$$\begin{aligned}
\widehat{\mathbf{b}} &= \widehat{\boldsymbol{\Sigma}}^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2) \\
\widehat{b}_0 &= -\frac{1}{2}\left[\overline{\mathbf{x}}_1^T\widehat{\boldsymbol{\Sigma}}^{-1}\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2^T\widehat{\boldsymbol{\Sigma}}^{-1}\overline{\mathbf{x}}_2\right] + \log_e\frac{n_1}{n} - \log_e\frac{n_2}{n}
\end{aligned}$$

are ML estimates of $\mathbf{b}$ and $b_0$ respectively.

# LDA

The probabilities $\pi_1, \pi_2$ can be chosen based on past experiences or can be estimated as

$$\widehat{\pi}_i = \frac{n_i}{n}, i = 1, 2$$

Then $\widehat{L}(\mathbf{x}) = \widehat{b}_0 + \widehat{\mathbf{b}}^T\mathbf{x}$, where

$$\begin{aligned} \widehat{\mathbf{b}} &= \widehat{\boldsymbol{\Sigma}}^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2) \\ \widehat{b}_0 &= -\frac{1}{2}\left[\overline{\mathbf{x}}_1^T\widehat{\boldsymbol{\Sigma}}^{-1}\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2^T\widehat{\boldsymbol{\Sigma}}^{-1}\overline{\mathbf{x}}_2\right] + \log_e\frac{n_1}{n} - \log_e\frac{n_2}{n} \end{aligned}$$

are ML estimates of $\mathbf{b}$ and $b_0$ respectively.

Classification rule The classification rule assigns $\mathbf{x}$ to $\Pi_1$ if $\widehat{L}(\mathbf{x}) > 0$, and assigns $\mathbf{x}$ to $\Pi_2$ otherwise.

# Quadratic Discriminant Analysis

Question How would the classification be affected if the covariance matrices of the two Gaussian populations are not equal to each other?

# Quadratic Discriminant Analysis

Question How would the classification be affected if the covariance matrices of the two Gaussian populations are not equal to each other?

Then

$$\begin{aligned}\log_e \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &= c_0 - \frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\mathbf{x}-\boldsymbol{\mu}_1) - (\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}_2^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)\right] \\ &= c_1 - \frac{1}{2}\mathbf{x}^T(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1})\mathbf{x} + (\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\mu}_2^T\boldsymbol{\Sigma}_2^{-1})\mathbf{x},\end{aligned}$$

where $c_0$ and $c_1$ are constants that depend on $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$.

# QDA

Thus the log-likelihood ration has the form of a quadratic function of $\mathbf{x}$ :

$$Q(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^T \mathbf{x} + \mathbf{x}^T \boldsymbol{\Omega} \mathbf{x},$$

where

$$\begin{aligned}
\boldsymbol{\Omega} &= -\frac{1}{2}(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1}) \\
\boldsymbol{\beta} &= \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 \\
\beta_0 &= -\frac{1}{2}\left[\log_e \frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|} + \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right] - \log_e(\pi_2/\pi_1)
\end{aligned}$$

# QDA

Thus the log-likelihood ration has the form of a quadratic function of $\mathbf{x}$ :

$$Q(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^T \mathbf{x} + \mathbf{x}^T \boldsymbol{\Omega} \mathbf{x},$$

where

$$\begin{aligned}
\boldsymbol{\Omega} &= -\frac{1}{2}(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1}) \\
\boldsymbol{\beta} &= \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 \\
\beta_0 &= -\frac{1}{2}\left[\log_e \frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|} + \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right] - \log_e(\pi_2/\pi_1)
\end{aligned}$$

Classification rule If $Q(\mathbf{x}) > 0$ assign $\mathbf{x}$ to $\Pi_1$, and assign $\mathbf{x}$ to $\Pi_2$ otherwise.

# QDA

Thus the log-likelihood ration has the form of a quadratic function of $\mathbf{x}$ :

$$Q(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^T \mathbf{x} + \mathbf{x}^T \boldsymbol{\Omega} \mathbf{x},$$

where

$$\begin{aligned}
\boldsymbol{\Omega} &= -\frac{1}{2}(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1}) \\
\boldsymbol{\beta} &= \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 \\
\beta_0 &= -\frac{1}{2}\left[\log_e \frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|} + \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right] - \log_e(\pi_2/\pi_1)
\end{aligned}$$

Classification rule If $Q(\mathbf{x}) > 0$ assign $\mathbf{x}$ to $\Pi_1$, and assign $\mathbf{x}$ to $\Pi_2$ otherwise.

Question How do you implement it in real data?