STAT 380:
Classification Technique: Discriminant Analysis

UAEU

Outlook of the unit

- **Prediction and Classification Approaches**
  - Classification Techniques
    - Logistic regression
    - Discriminant analysis
  - Evaluating Performance of a Classification Technique
  - Tree-based methods: Decision trees
    - Classification trees
    - Regression trees

Classification Technique: Discriminant Analysis

- **Discriminant Analysis** can also model the response variable that contains **more than two categories.**

- It formulates the probability of a data point belongs to a given category based on the corresponding values of the explanatory variables/covariates.

- Linear Discriminant Analysis (LDA) and the Quadratic Discriminant Analysis (QDA) are often considered as the default option to model categorical variables that has more than two classes.

Assumptions

However, it is developed based on the following assumption: The distribution of the predictors X is **Normal** in each of the classes.

It employs the **Bayes Principle** to obtain the probability that a data-point belongs to each of response categories.

It predicts to assign a specific data-point to the response category that has the maximum probability to contain the point.

Assumptions

However, it is developed based on the following assumption: The distribution of the predictors X is **Normal** in each of the classes.

It employs the **Bayes Principle** to obtain the probability that a data-point belongs to each of response categories.

It predicts to assign a specific data-point to the response category that has the maximum probability to contain the point.

Review: Normal Distribution

Normal Distribution

Notation:

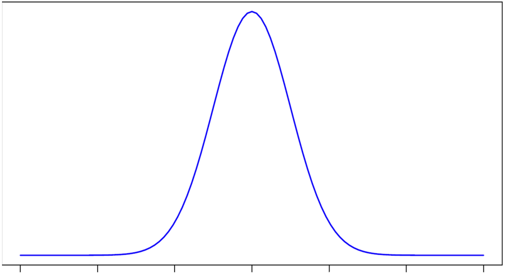
$\text{Normal}(\mu, \sigma)$

$-\infty < \mu < \infty$

$\sigma > 0$

μ: Mean Parameter

σ: Standard Deviation



Probability Density Function: $\phi(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

Normal Distribution

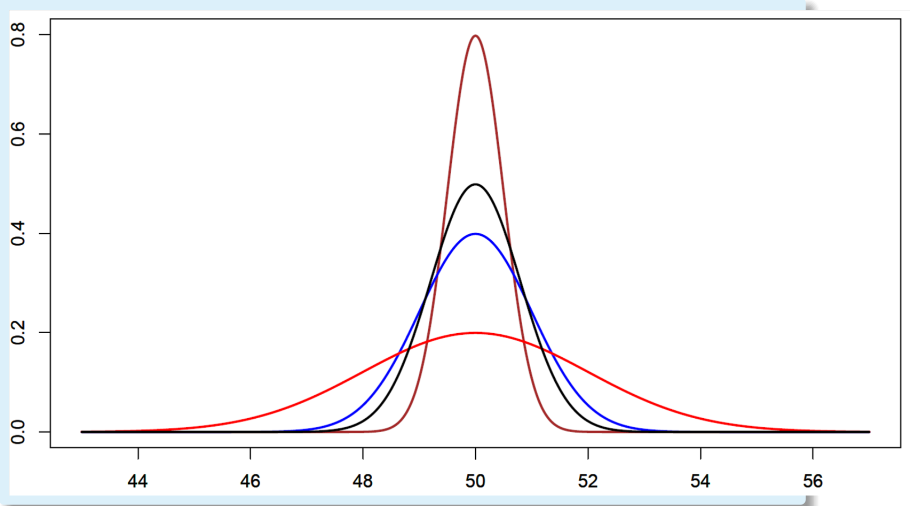
- The distribution is symmetric about its mean μ .

The mean is also called the **location** parameter.

The standard deviation corresponds to the **scale** of the distribution..

Effect on the Normal density if μ is changed

Effect on the Normal density if σ is changed

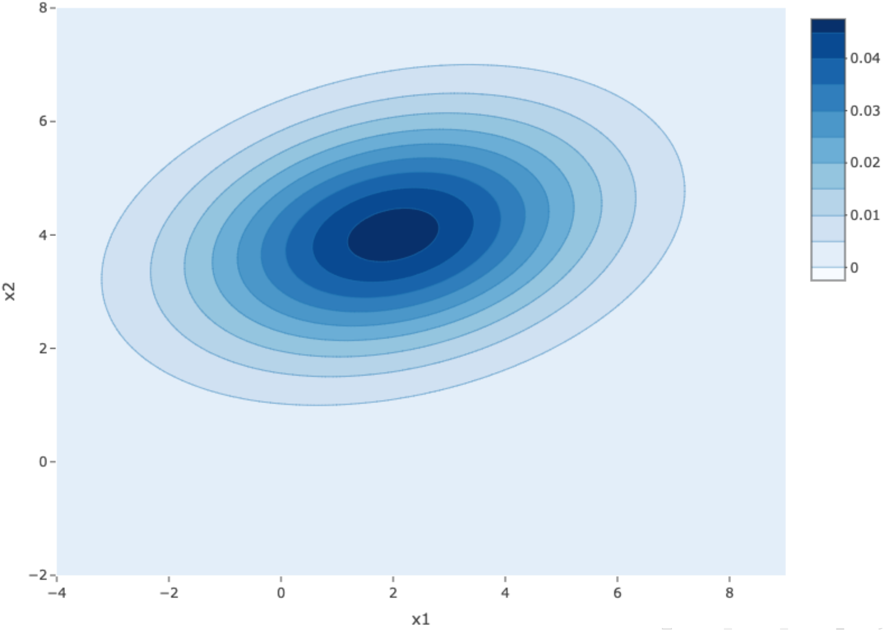


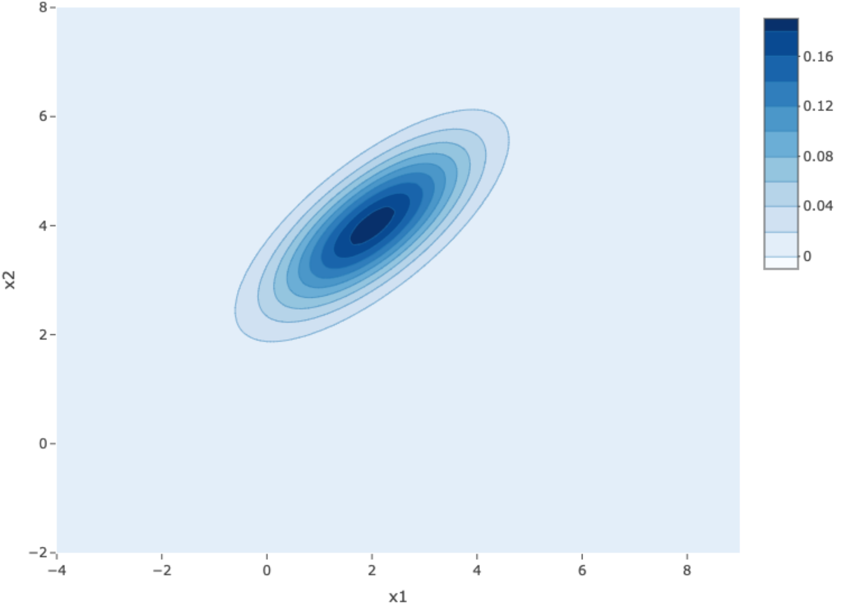
Review: Multivariate Normal Distribution

Normal Distribution

Probability Density Function:

$$\phi_p(\underset{\sim}{\mathbf{x}} \mid \underset{\sim}{\mu}, \Sigma) = \frac{1}{\sqrt{|\Sigma|}\left(\sqrt{2\pi}\right)^{\frac{p}{2}}} \exp\left(-(\underset{\sim}{\mathbf{x}} - \underset{\sim}{\mu})^T \Sigma^{-1} (x - \underset{\sim}{\mu})\right)$$


8
6
4
2
0
−2
−4
−2
0
2
4
6
8
x1
x2
0.04
0.03
0.02
0.01
0



8
6
4
2
0
−2
x2
−4
−2
0
2
4
6
8
x1
0.16
0.12
0.08
0.04
0

Review: Parameter Estimate for Normal Distribution

$Z_1, Z_2, \ldots, Z_m \overset{i.i.d.}{\sim}$ Normal($\underset{\sim}{\mu}, \Sigma$) then

$$\widehat{\underset{\sim}{\mu}} := \frac{\sum_{i=1}^{m} Z_i}{m} \quad \text{and} \quad \widehat{\Sigma} := \frac{\sum_{i=1}^{m} \left(Z_i - \widehat{\underset{\sim}{\mu}}\right)\left(Z_i - \widehat{\underset{\sim}{\mu}}\right)^T}{m-1}$$

Review: Bayes Principle

Law of Conditional Probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \text{ given } B) = \frac{P(A, \text{and } B)}{P(B)}$$

$$P(A \mid B)P(B) = P(A \cap B)$$

Law of Conditional Probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \text{ given } B) = \frac{P(A, \text{and } B)}{P(B)}$$

$$P(A \mid B)P(B) = P(A \cap B)$$

Bayes Rule

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

Bayes Rule: In the Current Context

$$\boxed{P(Y = g \mid X)} = \frac{\boxed{P(X \mid Y = g)} \times P(Y = g)}{P(X)}$$

General Form: Decision Rule in Discriminant Analysis

To decide whether a data point with covariate value x_i belongs to a group among $\{1, 2, \ldots, G\}$, A decission score of the following form is calculated: $\delta_g(x_i)$ for $g = 1, \ldots, G$.

$\delta_g(x_i)$:Score that x_i belongs to Group g

The point is Assigned to the group with maximum δ score.

General Form: Decision Rule in Discriminant Analysis

For Example: If there is only two groups; *Group*1 and *Group*2. Then a point with covariate value x_i is:

Assigned to the group $g = 1$ if $\delta_1(x_i) > \delta_2(x_i)$

Assigned to the group $g = 2$ if $\delta_2(x_i) \geq \delta_1(x_i)$

General Form: Decision Rule in Discriminant Analysis

Based on the different assumptions, QDA and LDA results in different decission score function δ.

Both the procedures are calculated from the formula: $\log(P(Y = g \mid X))$.

Although used the same strategy, the differences in the decession score δ that we see for the two procedures is because of the differentces in their corresponding assumptions.

We will see that:

The δ function for LDA is **Linear function** of the covariates, x.

The δ function for QDA is **Quadratic function** of the covariates, x.

Details of LDA

LDA: Modelling Assumptions

- Data: $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ where $x_i \in \mathbb{R}$ and $y_i \in \{1, 2, \ldots, G\}$

$$x_i \mid y_i = 1 \sim \text{Normal}\left(\mu_1, \sigma^2\right)$$

$$x_i \mid y_i = 2 \sim \text{Normal}\left(\mu_2, \sigma^2\right)$$

$$\vdots$$

$$x_i \mid y_i = G \sim \text{Normal}\left(\mu_G, \sigma^2\right)$$

Posterior Probability of a Point Belongs to a Group

$$
\begin{aligned}
P(Y_i = \boxed{g} \mid X = x_i) &= \frac{\boxed{P(x_i \mid Y_i = \boxed{g})} \times P(Y_i = \boxed{g})}{P(x_i)} \\
&= \frac{\boxed{\phi(x_i \mid \boxed{\mu_g}, \boxed{\sigma^2})} \times \pi_g}{P(x_i)}
\end{aligned}
$$

Decision Rule if We have **Two** groups

$$\frac{P(Y_i = \boxed{1} \mid X = x_i)}{P(Y_i = \boxed{2} \mid X = x_i)} = \frac{\phi(x_i \mid \boxed{\mu_1}, \boxed{\sigma^2}) \times \pi_1}{\phi(x_i \mid \boxed{\mu_2}, \boxed{\sigma^2}) \times \pi_2}$$

Assign the point to the group $g = 1$ if $\frac{P(Y_i = \boxed{1} \mid X = x_i)}{P(Y_i = \boxed{2} \mid X = x_i)} > 1$

Assign the point to the group $g = 2$ if $\frac{P(Y_i = \boxed{1} \mid X = x_i)}{P(Y_i = \boxed{2} \mid X = x_i)} \leq 1$

Decision Rule if We have **Two** groups

$$\frac{P(Y_i = \boxed{1} \mid X = x_i)}{P(Y_i = \boxed{2} \mid X = x_i)} = \frac{\phi(x_i \mid \boxed{\mu_1}, \boxed{\sigma^2}) \times \pi_1}{\phi(x_i \mid \boxed{\mu_2}, \boxed{\sigma^2}) \times \pi_2}$$

Assign the point to the group $g = 1$ if $\frac{P(Y_i = \boxed{1} | X = x_i)}{P(Y_i = \boxed{2} | X = x_i)} > 1$

Assign the point to the group $g = 2$ if $\frac{P(Y_i = \boxed{1} | X = x_i)}{P(Y_i = \boxed{2} | X = x_i)} \leq 1$

Decision Rule if We have **Two** groups

$$\log\left(\frac{P(Y_i = \boxed{1}\,|X = x_i)}{P(Y_i = \boxed{2}\,|X = x_i)}\right) = \boxed{\frac{(\mu_1 - \mu_2)}{\sigma^2}\, x} + \log\left(\frac{\pi_1}{\pi_2}\right) - \frac{\mu_1^1 - \mu_2^2}{2\sigma^2}$$

Decision Rule if We have **Two** groups

$$
\begin{aligned}
&\log\left(\frac{P(Y_i = 1 \mid X = x_i)}{P(Y_i = 2 \mid X = x_i)}\right) \\
&= \frac{(\mu_1 - \mu_2)}{\sigma^2}\, x + \log\left(\frac{\pi_1}{\pi_2}\right) - \frac{\mu_1^1 - \mu_2^2}{2\sigma^2} \\
&= \left[\frac{\mu_1}{\sigma^2} x + \log(\pi_1) - \frac{\mu_1^2}{2\sigma^2}\right] - \left[\frac{\mu_2}{\sigma^2} x + \log(\pi_2) - \frac{\mu_2^2}{2\sigma^2}\right] \\
&= \delta_1 - \delta_2
\end{aligned}
$$

Estimates of the Parameters from the training Sample

Finally: LDA- Decision Rule

- $\delta_g(x) = \log(\widehat{\pi}_g) + \frac{\widehat{\mu}_g}{\widehat{\sigma}^2} x - \frac{\widehat{\mu}_g^2}{2\widehat{\sigma}^2}$ for $g = 1, 2$.

- Then a point with covariate value x_i is:
  - Assigned to the group $g = 1$ if $\delta_1(x_i) > \delta_2(x_i)$
  - Assigned to the group $g = 2$ if $\delta_2(x_i) \geq \delta_1(x_i)$

Finally: LDA- Decision Rule

- $\delta_g(x) = \log(\widehat{\pi}_g) + \frac{\widehat{\mu}_g}{\widehat{\sigma}^2}x - \frac{\widehat{\mu}_g^2}{2\widehat{\sigma}^2}$ for $g = 1, 2$.

- Then a point with covariate value x_i is:

  Assigned to the group $g = 1$ if $\delta_1(x_i) > \delta_2(x_i)$

  Assigned to the group $g = 2$ if $\delta_2(x_i) \geq \delta_1(x_i)$

Multivariate LDA: Modelling Assumptions

- Data: $\{(\underset{\sim}{\mathbf{x}}_1, y_1), (\underset{\sim}{\mathbf{x}}_2, y_2), \ldots, (\underset{\sim}{\mathbf{x}}_n, y_n)\}$ where $\underset{\sim}{\mathbf{x}}_i \in \mathbb{R}$ and $y_i \in \{1, 2, \ldots, G\}$

$$\underset{\sim}{\mathbf{x}}_i \mid y_i = 1 \sim \text{Normal}\left(\underset{\sim}{\mu}_1, \Sigma\right)$$

$$\underset{\sim}{\mathbf{x}}_i \mid y_i = 2 \sim \text{Normal}\left(\underset{\sim}{\mu}_2, \Sigma\right)$$

$$\vdots$$

$$\underset{\sim}{\mathbf{x}}_i \mid y_i = G \sim \text{Normal}\left(\underset{\sim}{\mu}_G, \Sigma\right)$$

- $\delta_g(\underset{\sim}{\mathbf{x}}) = \log(\widehat{\pi_g}) + \underset{\sim}{\widehat{\mu}}_g^T \widehat{\Sigma}^{-1} \underset{\sim}{\mathbf{x}} - \frac{1}{2} \underset{\sim}{\widehat{\mu}}_g^T \Sigma^{-1} \underset{\sim}{\widehat{\mu}}_g$ for $g = 1, 2, \ldots, G$.

- Then a point with covariate $\underset{\sim}{\mathbf{x}}$ is asigned to the group $g^\star$ if :

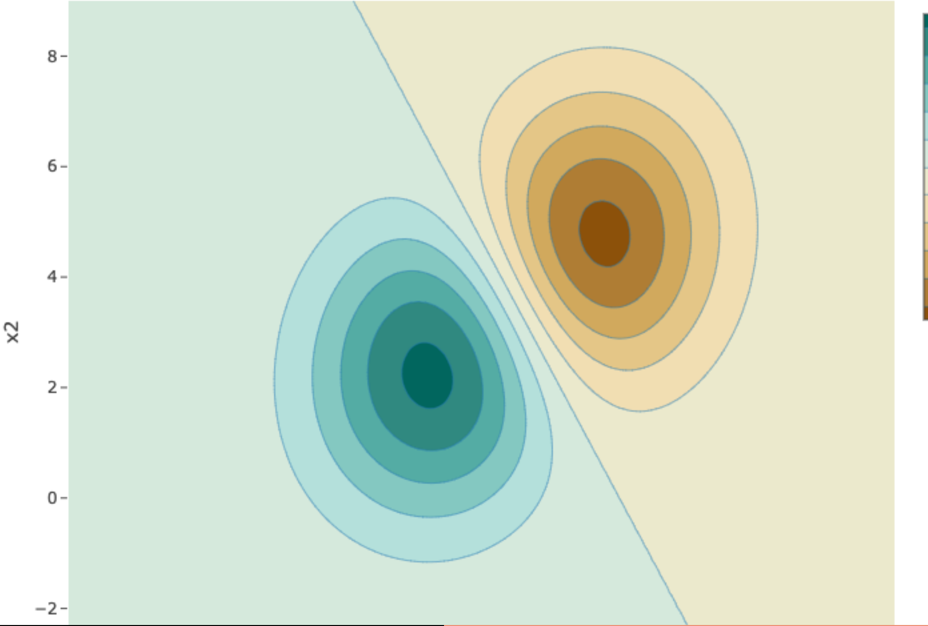
$$g^\star = \underset{g \in 1, \ldots, G}{\text{Arg max}} \left\{ \delta_g(\underset{\sim}{\mathbf{x}}) \right\}$$

Multivariate LDA: Decision Rule G groups, $G \geq 2$

$$\delta_g(\underset{\sim}{\mathbf{x}}) = \log(\widehat{\pi_g}) + \widehat{\underset{\sim}{\mu}}_g^T \widehat{\Sigma}^{-1} \underset{\sim}{\mathbf{x}} - \frac{1}{2} \widehat{\underset{\sim}{\mu}}_g^T \Sigma^{-1} \widehat{\underset{\sim}{\mu}}_g \quad \text{for } g = 1, 2, \ldots, G.$$

Then a point with covariate $\underset{\sim}{\mathbf{x}}$ is asigned to the group $g^\star$ if :

$$g^\star = \underset{g \in 1,\ldots,G}{\text{Arg max}} \left\{ \delta_g(\underset{\sim}{\mathbf{x}}) \right\}$$



8
6
4
2
0
−2
x2

Details of QDA

QDA: Modelling Assumptions

Data: $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ where $x_i \in \mathbb{R}$ and $y_i \in \{1, 2, \ldots, G\}$

$$x_i \mid y_i = 1 \sim \text{Normal}\left(\mu_1, \sigma_1^2\right)$$

$$x_i \mid y_i = 2 \sim \text{Normal}\left(\mu_2, \sigma_2^2\right)$$

$$\vdots$$

$$x_i \mid y_i = G \sim \text{Normal}\left(\mu_G, \sigma_G^2\right)$$

Posterior Probability of a Point Belongs to a Group

$$
\begin{aligned}
P(Y_i = g \mid X = x_i) &= \frac{P(x_i \mid Y_i = g) \times P(Y_i = g)}{P(x_i)} \\
&= \frac{\phi(x_i \mid \mu_g, \sigma_g^2) \times \pi_g}{P(x_i)}
\end{aligned}
$$

Decision Rule if We have **Two** groups

$$\log\left(\frac{P(Y_i=1|X=x_i)}{P(Y_i=2|X=x_i)}\right) = \delta_1 - \delta_2$$

$$\delta_g = \log(\pi_g) - \frac{\log(\sigma_g^2)}{2} - \frac{(x-\mu_g)^2}{2\sigma_g^2}$$ for $g = 1, 2$.

- Then a point with covariate value x_i is:

Assigned to the group $g = 1$ if $\delta_1(x_i) > \delta_2(x_i)$

Assigned to the group $g = 2$ if $\delta_2(x_i) \geq \delta_1(x_i)$

Multivariate QDA: Modelling Assumptions

Data: $\{(\underset{\sim}{\mathbf{x}}_1, y_1), (\underset{\sim}{\mathbf{x}}_2, y_2), \ldots, (\underset{\sim}{\mathbf{x}}_n, y_n)\}$ where $\underset{\sim}{\mathbf{x}}_i \in \mathbb{R}$ and $y_i \in \{1, 2, \ldots, G\}$

$$\underset{\sim}{\mathbf{x}}_i \mid y_i = 1 \sim \text{Normal}\left(\underset{\sim}{\mu}_1, \Sigma_1\right)$$

$$\underset{\sim}{\mathbf{x}}_i \mid y_i = 2 \sim \text{Normal}\left(\underset{\sim}{\mu}_2, \Sigma_2\right)$$

$$\vdots$$

$$\underset{\sim}{\mathbf{x}}_i \mid y_i = G \sim \text{Normal}\left(\underset{\sim}{\mu}_G, \Sigma_G\right)$$

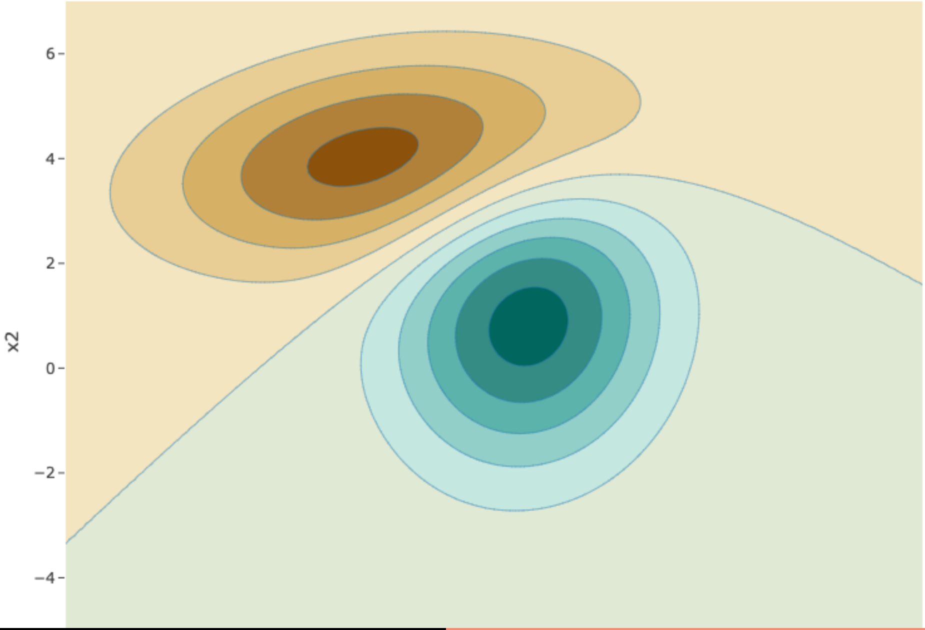
Multivariate LDA: Decision Rule G groups, $G \geq 2$

$$\delta_g(\underset{\sim}{\mathbf{x}}) = \log(\pi_g) - \frac{1}{2}\log\left(|\Sigma_g|\right) - \left(\underset{\sim}{\mathbf{x}} - \underset{\sim}{\mu}_g\right)^T \Sigma^{-1} \left(\underset{\sim}{\mathbf{x}} - \underset{\sim}{\mu}_g\right)$$

for $g = 1, 2, \ldots, G$.

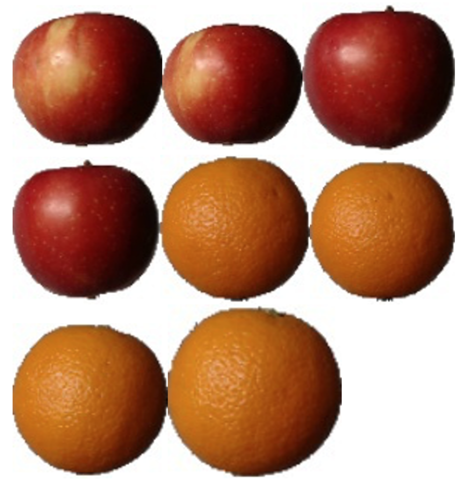
- Then a point with covariate value $\underset{\sim}{\mathbf{x}}$ is asigned to k if :

$$k = \underset{g \in 1,\ldots,G}{\text{Arg max}} \left\{ \delta_g(\underset{\sim}{\mathbf{x}}) \right\}$$


6
4
2
0
−2
−4
x2

Example

Example: "Apples" from the "Oranges" Dataset



How about this?