

Foundations of Machine Learning

Classification: decisions and discriminants

<https://eduassistpro.github.io/>

Srinandan (“Sri”) Dasmahapatra
[Add WeChat edu_assist_pro](#)

2-class classification: determine which class a given input
 $X=x$ belongs to

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

2-class classification: determine which class a given input $X=x$ belongs to

decision boundary at equal posterior probability for either class, which are:

$$P(C = +|X = x) = \frac{P(X = x|C = +)P(C = +)}{P(X = x|C = +)P(C = +) + P(X = x|C = -)P(C = -)}$$
$$P(C = -|X = x) = \frac{P(X = x|C = -)P(C = -)}{P(X = x|C = +)P(C = +) + P(X = x|C = -)P(C = -)}$$

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

2-class classification: determine which class a given input $X=x$ belongs to

decision boundary at equal posterior probability for either class, which are:

$$P(C = +|X = x) = \frac{P(X = x|C = +)P(C = +)}{P(X = x|C = +)P(C = +) + P(X = x|C = -)P(C = -)}$$

$$P(C = -|X = x) = \frac{P(X = x|C = -)P(C = -)}{P(X = x|C = +)P(C = +) + P(X = x|C = -)P(C = -)}$$

<https://eduassistpro.github.io/>

Simplify the notation: introduce [Add WeChat](#) `edu_assist_pro`

2-class classification: determine which class a given input $X=x$ belongs to

decision boundary at equal posterior probability for either class, which are:

$$P(C = +|X = x) = \frac{P(X = x|C = +)P(C = +)}{P(X = x|C = +)P(C = +) + P(X = x|C = -)P(C = -)}$$
$$P(C = -|X = x) = \frac{P(X = x|C = -)P(C = -)}{P(X = x|C = +)P(C = +) + P(X = x|C = -)P(C = -)}$$

<https://eduassistpro.github.io/>

Simplify the notation: introduce [Add WeChat](#) [edu_assist_pro](#)

$$P(C = +|X = x) = \frac{e^{g_+(x)}}{e^{g_+(x)} + e^{g_-(x)}} = \frac{\exp [g_+(x) - g_-(x)]}{1 + \exp [g_+(x) - g_-(x)]}$$
$$P(C = -|X = x) = \frac{e^{g_-(x)}}{e^{g_+(x)} + e^{g_-(x)}} = \frac{1}{1 + \exp [g_+(x) - g_-(x)]}$$

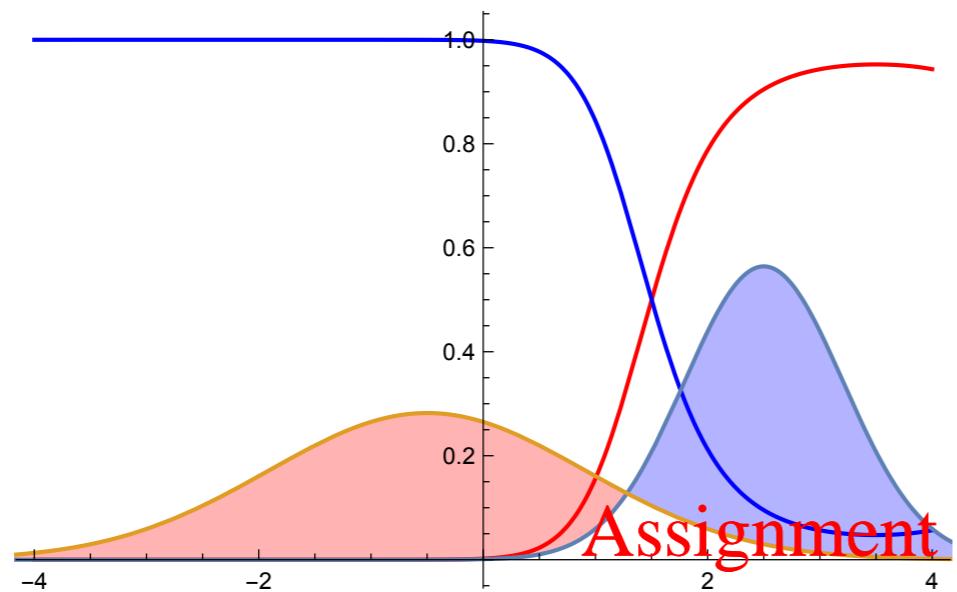
Decision boundary: determine which class a given input belongs to

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Decision boundary: determine which class a given input belongs to



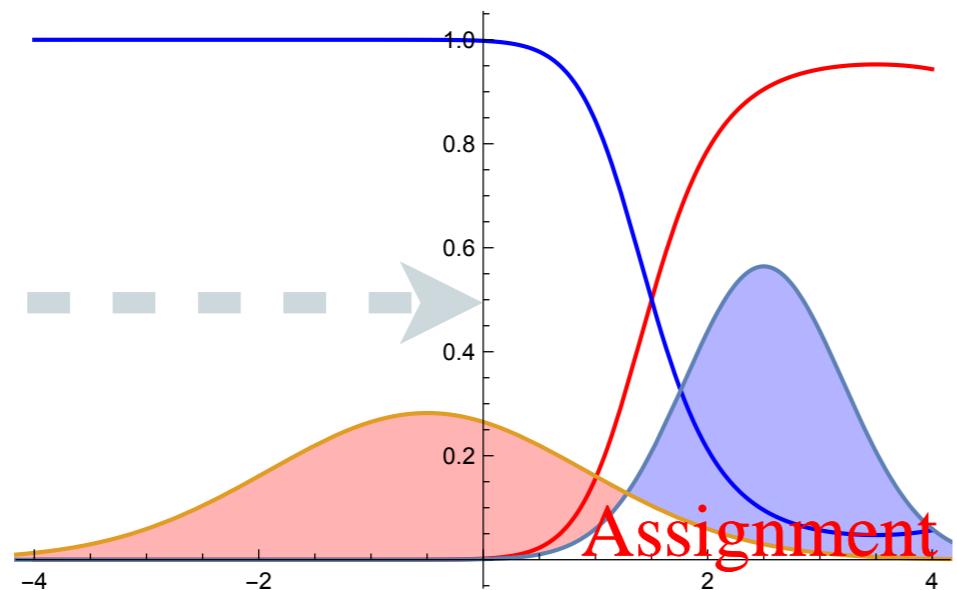
decision boundary at equal posterior probability (plotted) for either class

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Decision boundary: determine which class a given input belongs to



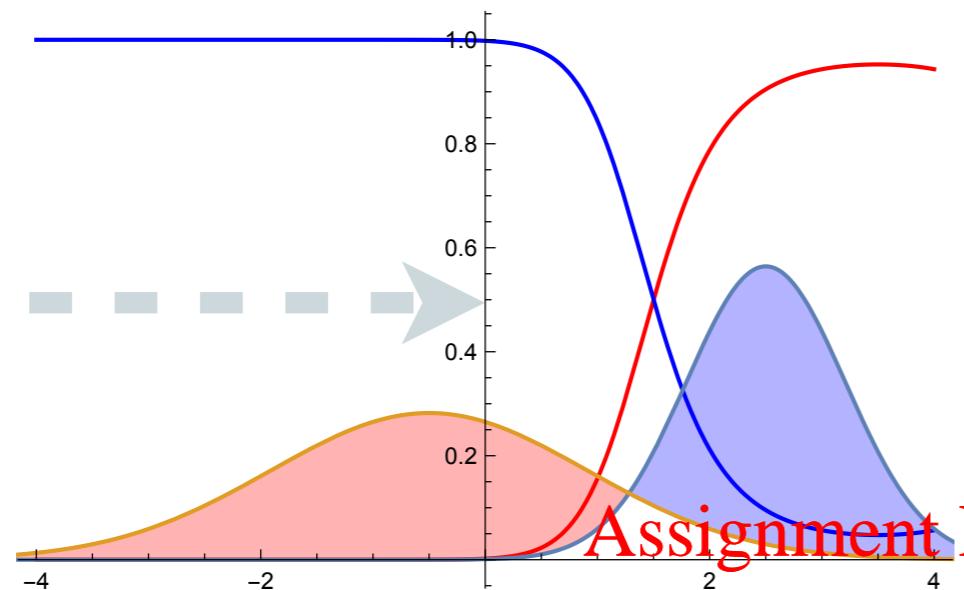
decision boundary at equal posterior probability (plotted) for either class

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Decision boundary: determine which class a given input belongs to



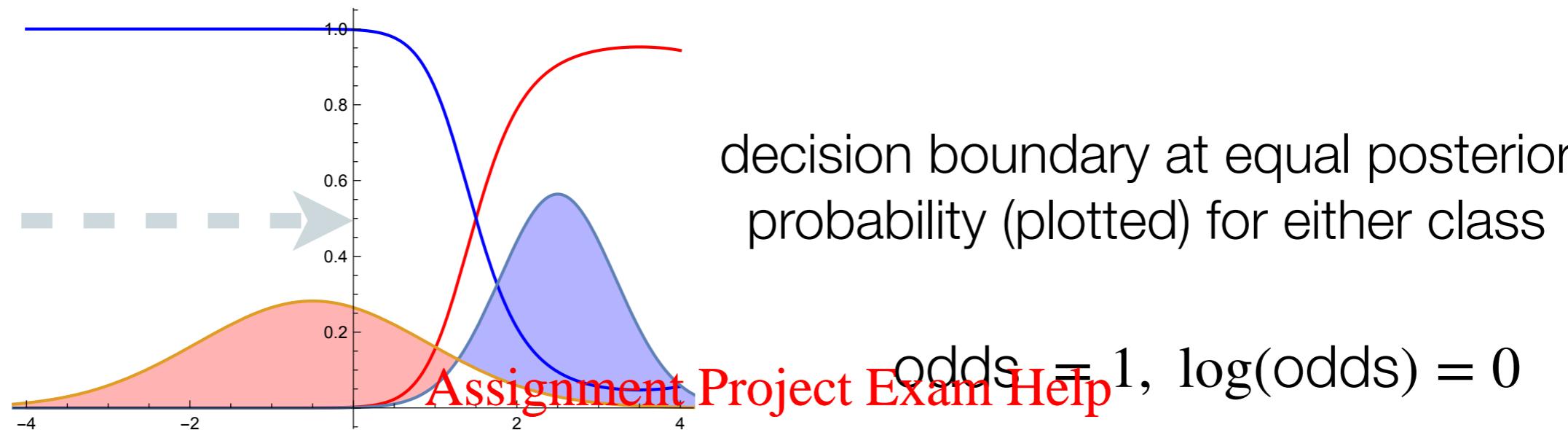
decision boundary at equal posterior probability (plotted) for either class

odds = 1, $\log(\text{odds}) = 0$

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Decision boundary: determine which class a given input belongs to



$$\text{odds} = \frac{P(C_1)}{P(C_0)} = e^{g_+(x) - g_-(x)}$$

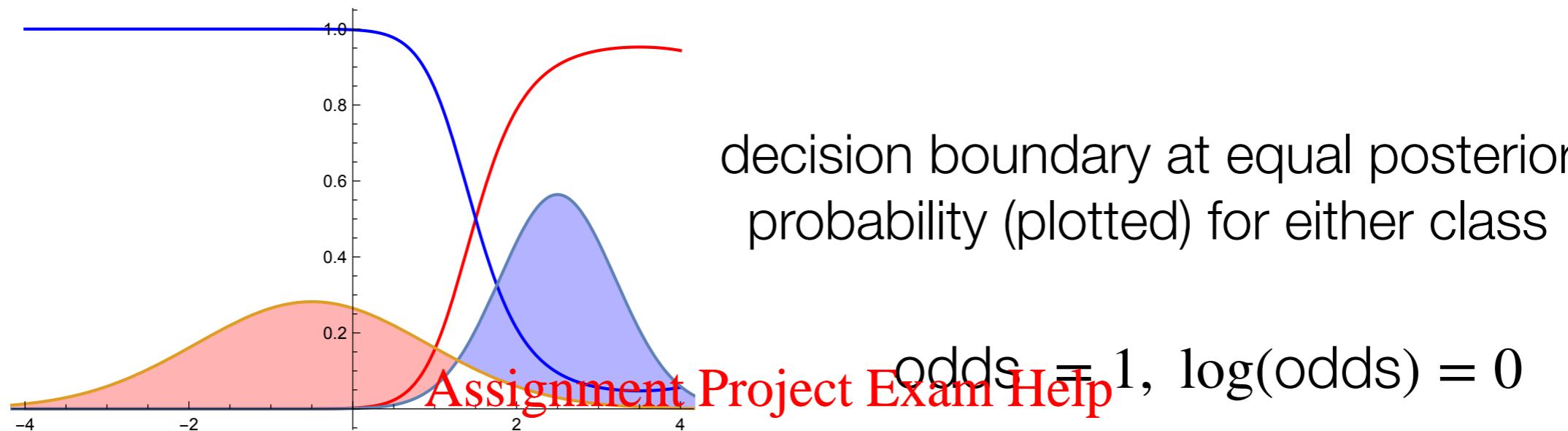
Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat

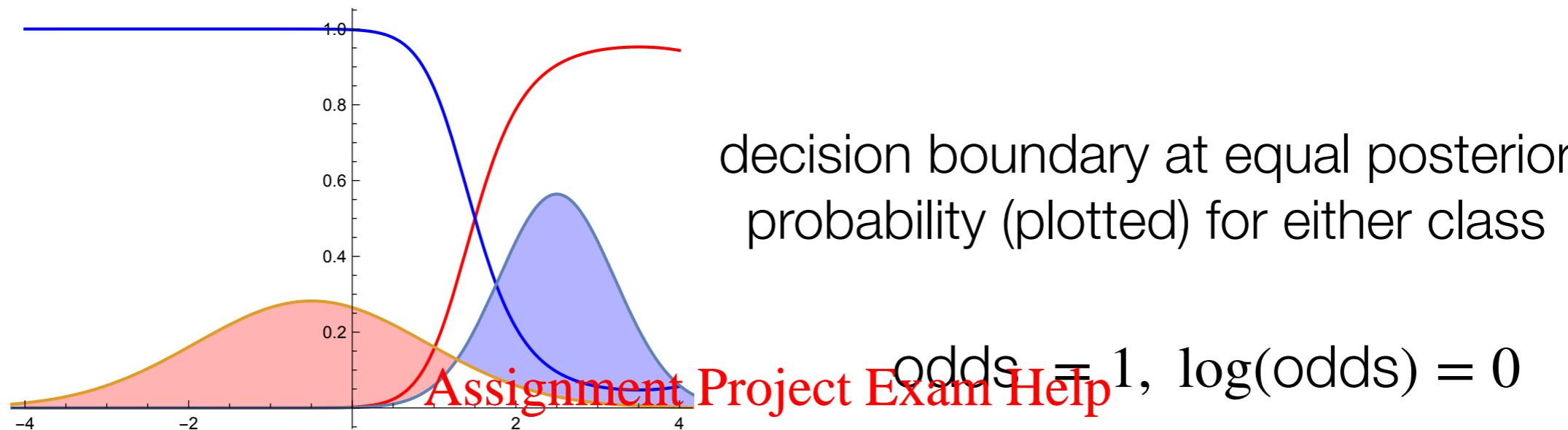
edu_assist_pro

Decision boundary: determine which class a given input belongs to



$$\text{odds} = \frac{P(C_1)}{P(C_0)} = e^{\beta_0 + \beta_1 x}$$

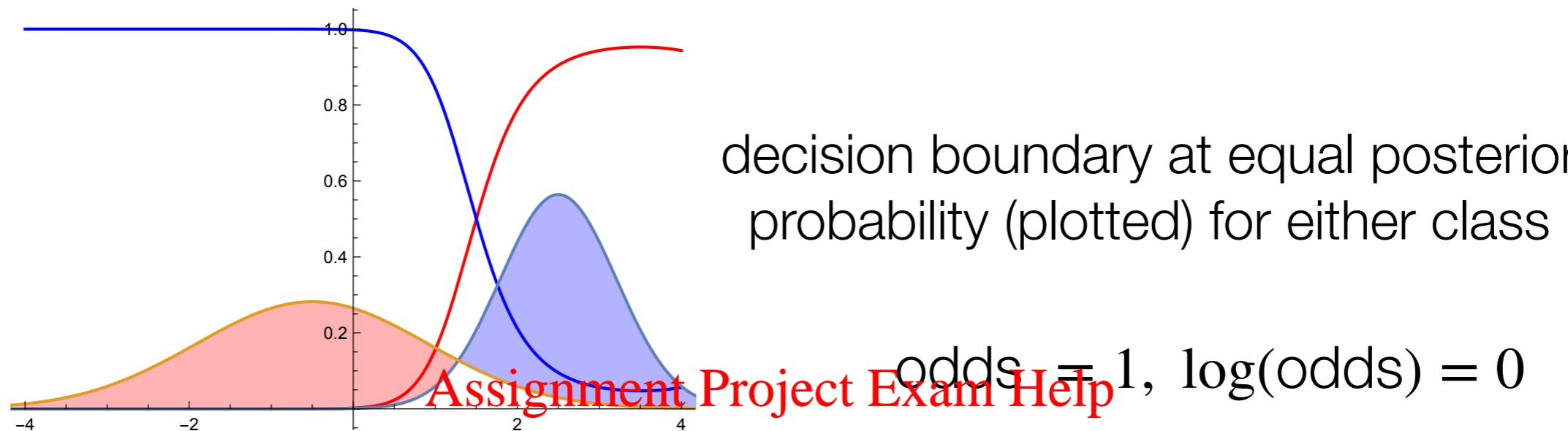
Decision boundary: determine which class a given input belongs to



$$\text{odds} = \frac{P(C \text{ https://eduassistpro.github.io/})}{P(C \text{ AddWeChat edu_assist_pro})} = e^{[g_+(x) - g_-(x)]}$$

$$\text{odds} = \frac{P(C = +|X = x)}{P(C = -|X = x)} = \frac{P(X = x|C = +)P(C = +)}{P(X = x|C = -)P(C = -)}$$

Decision boundary: determine which class a given input belongs to



$$\text{odds} = \frac{P(C \text{ https://eduassistpro.github.io/})}{P(C \text{ AddWeChat edu_assist_pro})} = p[g_+(x) - g_-(x)]$$

$$\text{odds} = \frac{P(C = +|X = x)}{P(C = -|X = x)} = \frac{P(X = x|C = +)P(C = +)}{P(X = x|C = -)P(C = -)}$$

$$\text{log-odds} = \ln \left(\frac{P(X = x|C = +)}{P(X = x|C = -)} \right) + \ln \left(\frac{P(C = +)}{P(C = -)} \right)$$

data dependent
(likelihood)

data independent
(prior)

Decision boundaries from probability distributions: zero
log-odds: $0 = g_+(\mathbf{x}) - g_-(\mathbf{x}) = f(\mathbf{x}; \theta) + \theta_0$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Decision boundaries from probability distributions: zero log-odds: $0 = g_+(\mathbf{x}) - g_-(\mathbf{x}) = f(\mathbf{x}; \theta) + \theta_0$

- If we choose linear functions:

$$g_+(\mathbf{x}) = w_{+0} + \mathbf{w}_+^\top \mathbf{x}, g_-(\mathbf{x}) = w_{-0} + \mathbf{w}_-^\top \mathbf{x}$$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Decision boundaries from probability distributions: zero log-odds: $0 = g_+(\mathbf{x}) - g_-(\mathbf{x}) = f(\mathbf{x}; \theta) + \theta_0$

- If we choose linear functions:

$$g_+(\mathbf{x}) = w_{+0} + \mathbf{w}_+^\top \mathbf{x}, \quad g_-(\mathbf{x}) = w_{-0} + \mathbf{w}_-^\top \mathbf{x}$$

- Then $f(\mathbf{x}; \theta) = \theta_0 + \theta^\top \mathbf{x}$ is a linear decision boundary

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Decision boundaries from probability distributions: zero log-odds: $0 = g_+(\mathbf{x}) - g_-(\mathbf{x}) = f(\mathbf{x}; \theta) + \theta_0$

- If we choose linear functions:

$$g_+(\mathbf{x}) = w_{+0} + \mathbf{w}_+^\top \mathbf{x}, \quad g_-(\mathbf{x}) = w_{-0} + \mathbf{w}_-^\top \mathbf{x}$$

- Then $f(\mathbf{x}; \theta) = \theta_0 + \theta^\top \mathbf{x}$ is a linear decision boundary
- where $\theta_0 = w_{+0} - \frac{\mathbf{w}_+^\top \mathbf{w}_-}{\mathbf{w}_+^\top \mathbf{w}_-}$
Add WeChat edu_assist_pro

Decision boundaries from probability distributions: zero log-odds: $0 = g_+(\mathbf{x}) - g_-(\mathbf{x}) = f(\mathbf{x}; \theta) + \theta_0$

- If we choose linear functions:

$$g_+(\mathbf{x}) = w_{+0} + \mathbf{w}_+^\top \mathbf{x}, \quad g_-(\mathbf{x}) = w_{-0} + \mathbf{w}_-^\top \mathbf{x}$$

- Then $f(\mathbf{x}; \theta) = \theta_0 + \theta^\top \mathbf{x}$ is a linear decision boundary
- where $\theta_0 = w_{+0} - \frac{\text{https://eduassistpro.github.io/}}{+} \mathbf{w}_-$
- This is 2-class softmax (logisti Add WeChat: edu_assist_pro ssion

Reminder: Logistic (2-class)/softmax (K-class) regression

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Reminder: Logistic (2-class)/softmax (K-class) regression

- Choose suitable features Φ to compute scores/activations:
 $s(A_i) := \text{score(class } A_i) = f_i(\Phi(\text{data}))$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Reminder: Logistic (2-class)/softmax (K-class) regression

- Choose suitable features Φ to compute scores/activations:

$$s(A_i) := \text{score(class } A_i) = f_i(\Phi(\text{data}))$$

- Recall,

Assignment Project Exam Help

$$\text{softmax}(s(A_1), \dots, s(A_K)) = \frac{\exp(s(A_1))}{\exp(s(A_1)) + \dots + \exp(s(A_K)))}$$

Add WeChat edu_assist_pro

Reminder: Logistic (2-class)/softmax (K-class) regression

- Choose suitable features Φ to compute scores/activations:
 $s(A_i) := \text{score(class } A_i) = f_i(\Phi(\text{data}))$
- Recall,
Assignment Project Exam Help
softmax($s(A_1), \dots, s(A_K)$)
$$\frac{\exp(s(A_1))}{\sum_{i=1}^K \exp(s(A_i))} + \dots + \frac{\exp(s(A_K))}{\sum_{i=1}^K \exp(s(A_i))}$$

<https://eduassistpro.github.io/>
- $P(\text{class } A_i | \text{data}) = \frac{\exp(s(A_i))}{\sum_{i=1}^K \exp(s(A_i))} = \frac{\exp(s(A_i))}{\sum_{i=1}^K \exp(s(A_i))}$
Add WeChat edu_assist_pro

Reminder: Logistic (2-class)/softmax (K-class) regression

- Choose suitable features Φ to compute scores/activations:
 $s(A_i) := \text{score(class } A_i) = f_i(\Phi(\text{data}))$
- Recall,
Assignment Project Exam Help
softmax($s(A_1), \dots, s(A_K)$)
$$\frac{\exp(s(A_1))}{\sum_{i=1}^K \exp(s(A_i))} + \dots + \frac{\exp(s(A_K))}{\sum_{i=1}^K \exp(s(A_i))}$$

<https://eduassistpro.github.io/>
- $P(\text{class } A_i | \text{data}) = \frac{\exp(s(A_i))}{\sum_{i=1}^K \exp(s(A_i))} = \frac{\exp(s(A_i))}{\sum_{i=1}^K \exp(s(A_i))}$
- $f_i(\text{data } x) = \mathbf{w}_i^\top \Phi(x) = \mathbf{w}_i^\top x$

Reminder: Logistic (2-class)/softmax (K-class) regression

- Choose suitable features Φ to compute scores/activations:
 $s(A_i) := \text{score(class } A_i) = f_i(\Phi(\text{data}))$
- Recall,
Assignment Project Exam Help
softmax($s(A_1), \dots, s(A_K)$)
$$\frac{\exp(s(A_1))}{\sum_{i=1}^K \exp(s(A_i))} + \dots + \frac{\exp(s(A_K))}{\sum_{i=1}^K \exp(s(A_i))}$$

<https://eduassistpro.github.io/>
- $P(\text{class } A_i | \text{data}) = \frac{\exp(s(A_i))}{\sum_{i=1}^K \exp(s(A_i))} = \frac{\exp(s(A_i))}{\sum_{i=1}^K \exp(s(A_i))}$
- $f_i(\text{data } x) = \mathbf{w}_i^\top \Phi(x) = \mathbf{w}_i^\top x$
- Did not seek a decision boundary, just an argmax

Decision boundaries from probability distributions: zero log-odds

Assignment Project Exam Help

<https://eduassistpro.github.io/>

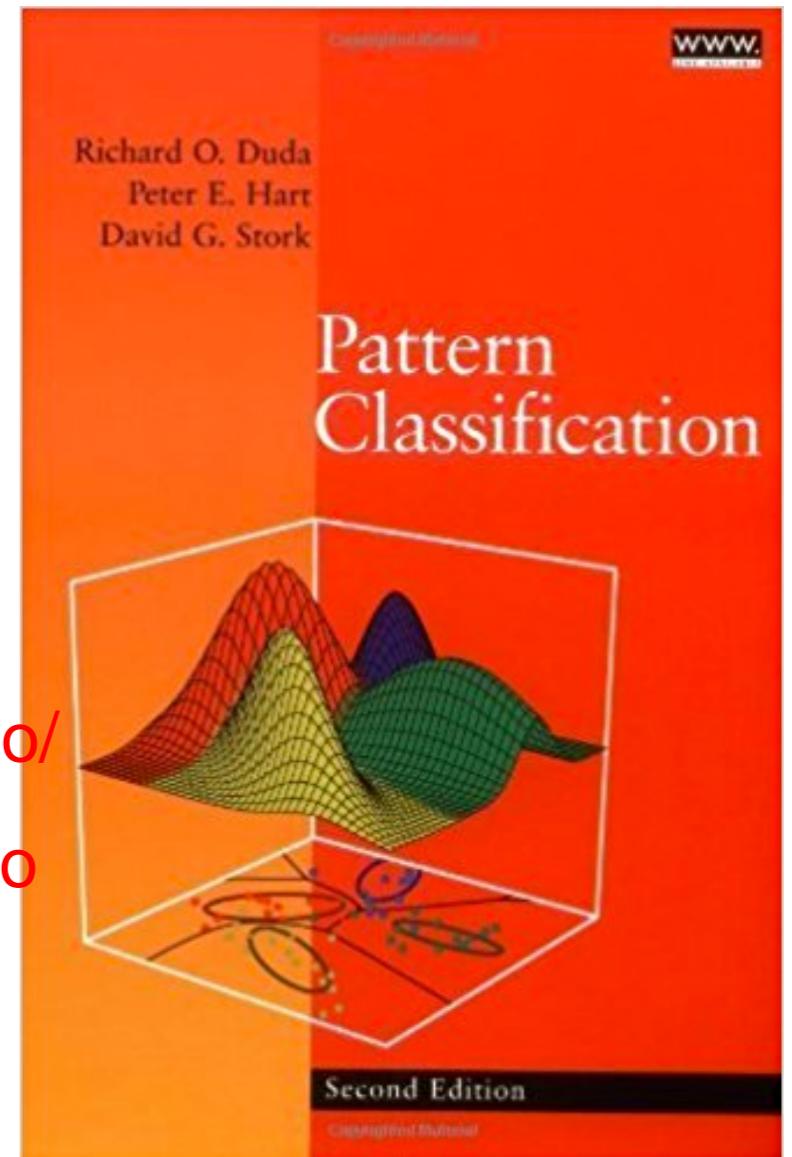
Add WeChat edu_assist_pro

Decision boundaries from probability distributions: zero log-odds

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

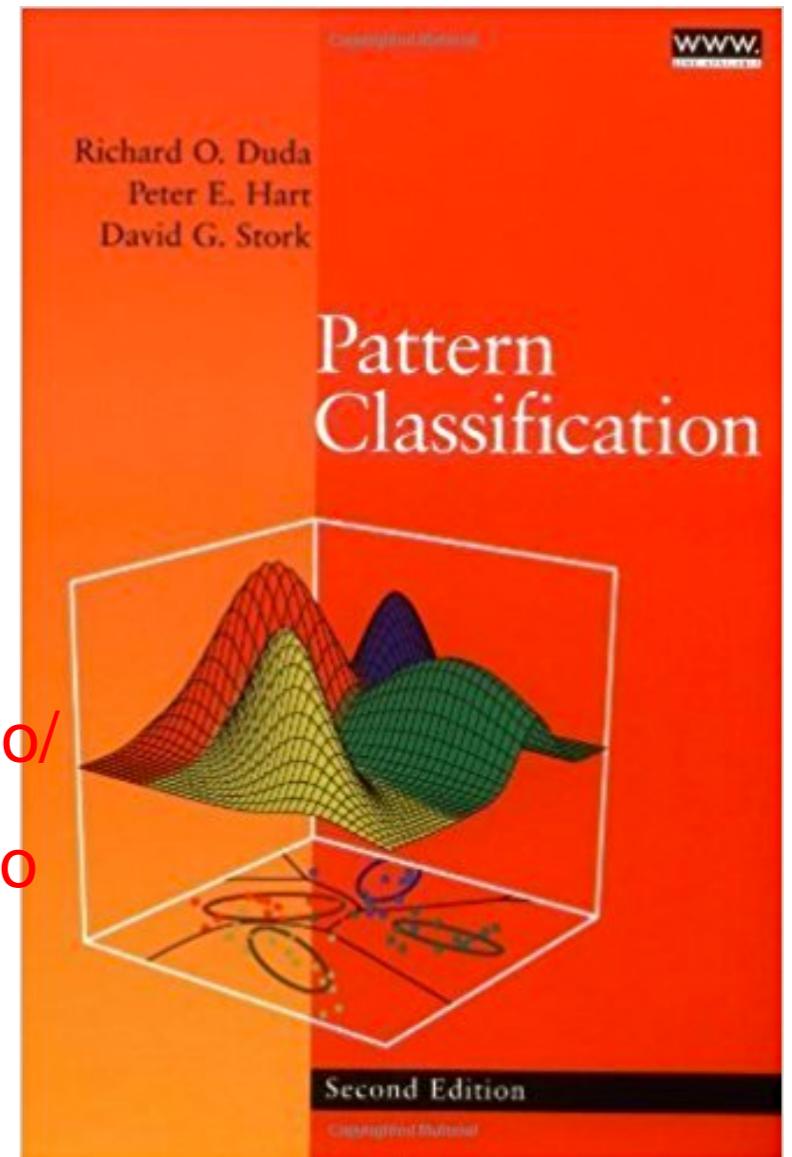


Decision boundaries from probability distributions: zero log-odds

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro



Notice the contours and
line segments at the
bottom of the cover graphic

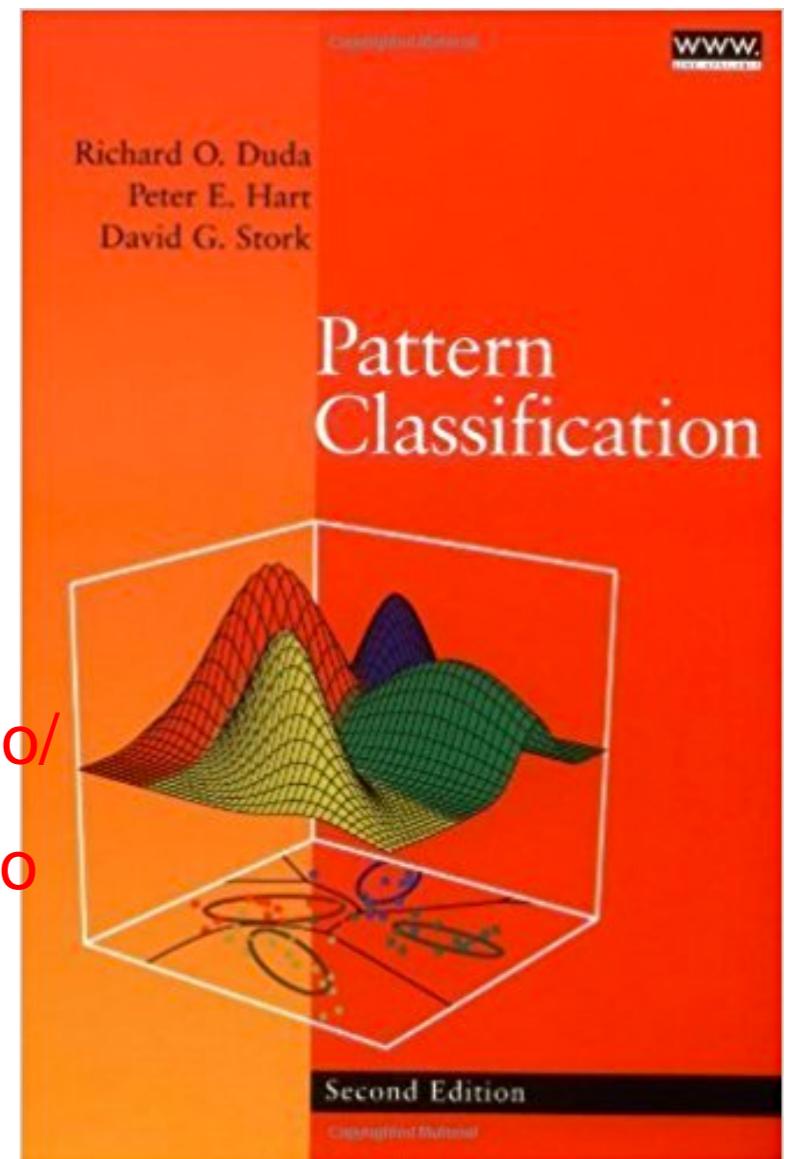
Decision boundaries from probability distributions: zero log-odds

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Bi-variate gaussian



Notice the contours and
line segments at the
bottom of the cover graphic

Log of ratio of two multi-dimensional Gaussians

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Log of ratio of two multi-dimensional Gaussians

$$p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Log of ratio of two multi-dimensional Gaussians

$$p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

Precision matrix $\Lambda = \boldsymbol{\Sigma}^{-1}$
Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Log of ratio of two multi-dimensional Gaussians

$$p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}|}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)$$

For 2D features

$$p(x_1, x_2 | C = a) = \frac{\sqrt{|\Lambda_a|}}{2\pi} \exp \left\{ -\frac{1}{2} \begin{bmatrix} x_1 - \mu_{a1} \\ x_2 - \mu_{a2} \end{bmatrix}^\top \begin{bmatrix} \Lambda_{a11} & \Lambda_{a12} \\ \Lambda_{a21} & \Lambda_{a22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_{a1} \\ x_2 - \mu_{a2} \end{bmatrix} \right\}$$

$$p(x_1, x_2 | C = b) = \frac{\sqrt{|\Lambda_b|}}{2\pi} \exp \left\{ -\frac{1}{2} \begin{bmatrix} x_1 - \mu_{b1} & x_2 - \mu_{b2} \end{bmatrix}^\top \begin{bmatrix} \Lambda_{b11} & \Lambda_{b12} \\ \Lambda_{b21} & \Lambda_{b22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_{b1} \\ x_2 - \mu_{b2} \end{bmatrix} \right\}$$

Precision matrix $\Lambda = \boldsymbol{\Sigma}^{-1}$
Assignment Project Exam Help

Add WeChat edu_assist_pro

<https://eduassistpro.github.io/>

Log of ratio of two multi-dimensional Gaussians

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

Precision matrix $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$
Assignment Project Exam Help

For 2D features

$$p(x_1, x_2 | C = a) = \frac{\sqrt{|\boldsymbol{\Lambda}_a|}}{2\pi} \exp \left\{ -\frac{1}{2} \begin{bmatrix} x_1 - \mu_{a1} \\ x_2 - \mu_{a2} \end{bmatrix}^T \begin{bmatrix} \Lambda_{a11} & \Lambda_{a12} \\ \Lambda_{a21} & \Lambda_{a22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_{a1} \\ x_2 - \mu_{a2} \end{bmatrix} \right\}$$

$$p(x_1, x_2 | C = b) = \frac{\sqrt{|\boldsymbol{\Lambda}_b|}}{2\pi} \exp \left\{ -\frac{1}{2} \begin{bmatrix} x_1 - \mu_{b1} & x_2 - \mu_{b2} \end{bmatrix}^T \begin{bmatrix} \Lambda_{b11} & \Lambda_{b12} \\ \Lambda_{b21} & \Lambda_{b22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_{b1} \\ x_2 - \mu_{b2} \end{bmatrix} \right\}$$

Log-odds (no prior) for 2-class classification

$$\ln \left(\frac{p(\mathbf{x}|C_+)}{p(\mathbf{x}|C_-)} \right) = \ln \left(\frac{|\boldsymbol{\Lambda}_+|}{|\boldsymbol{\Lambda}_-|} \right) + \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_-)^T \boldsymbol{\Lambda}_- (\mathbf{x} - \boldsymbol{\mu}_-) - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_+)^T \boldsymbol{\Lambda}_+ (\mathbf{x} - \boldsymbol{\mu}_+)$$

Log of ratio of posteriors of two Gaussians: for equal and unequal covariances

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Log of ratio of posteriors of two Gaussians: for equal and unequal covariances

$$\begin{aligned}g_-(\mathbf{x}) &= \ln P(C_-) + \frac{1}{2} \ln |\Lambda_-| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_-)^T \Lambda_- (\mathbf{x} - \boldsymbol{\mu}_-) \\&= \ln P(C_-) + \frac{1}{2} \ln |\Lambda_-| - \frac{1}{2} (\mathbf{x}^T \Lambda_- \mathbf{x} - 2\boldsymbol{\mu}_-^T \Lambda_- \mathbf{x} + \boldsymbol{\mu}_-^T \Lambda_- \boldsymbol{\mu}_-) \\g_+(\mathbf{x}) &= \ln P(C_+) + \frac{1}{2} \ln |\Lambda_+| - \frac{1}{2} (\mathbf{x}^T \Lambda_+ \mathbf{x} - 2\boldsymbol{\mu}_+^T \Lambda_+ \mathbf{x} + \boldsymbol{\mu}_+^T \Lambda_+ \boldsymbol{\mu}_+)\end{aligned}$$

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Log of ratio of posteriors of two Gaussians: for equal and unequal covariances

$$\begin{aligned}g_-(\mathbf{x}) &= \ln P(C_-) + \frac{1}{2} \ln |\Lambda_-| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_-)^T \Lambda_- (\mathbf{x} - \boldsymbol{\mu}_-) \\&= \ln P(C_-) + \frac{1}{2} \ln |\Lambda_-| - \frac{1}{2} (\mathbf{x}^T \Lambda_- \mathbf{x} - 2\boldsymbol{\mu}_-^T \Lambda_- \mathbf{x} + \boldsymbol{\mu}_-^T \Lambda_- \boldsymbol{\mu}_-) \\g_+(\mathbf{x}) &= \ln P(C_+) + \frac{1}{2} \ln |\Lambda_+| - \frac{1}{2} (\mathbf{x}^T \Lambda_+ \mathbf{x} - 2\boldsymbol{\mu}_+^T \Lambda_+ \mathbf{x} + \boldsymbol{\mu}_+^T \Lambda_+ \boldsymbol{\mu}_+)\end{aligned}$$

For unequal covariance <https://eduassistpro.github.io/>
Your turn: what is the quadratic form in \mathbf{x} for $g_+(\mathbf{x}) - g_-(\mathbf{x})$

Log of ratio of posteriors of two Gaussians: for equal and unequal covariances

$$\begin{aligned}g_-(\mathbf{x}) &= \ln P(C_-) + \frac{1}{2} \ln |\Lambda_-| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_-)^T \Lambda_- (\mathbf{x} - \boldsymbol{\mu}_-) \\&= \ln P(C_-) + \frac{1}{2} \ln |\Lambda_-| - \frac{1}{2} (\mathbf{x}^T \Lambda_- \mathbf{x} - 2\boldsymbol{\mu}_-^T \Lambda_- \mathbf{x} + \boldsymbol{\mu}_-^T \Lambda_- \boldsymbol{\mu}_-) \\g_+(\mathbf{x}) &= \ln P(C_+) + \frac{1}{2} \ln |\Lambda_+| - \frac{1}{2} (\mathbf{x}^T \Lambda_+ \mathbf{x} - 2\boldsymbol{\mu}_+^T \Lambda_+ \mathbf{x} + \boldsymbol{\mu}_+^T \Lambda_+ \boldsymbol{\mu}_+)\end{aligned}$$

For unequal covariance: <https://eduassistpro.github.io/>
Your turn: what is the quadratic function boundary? Add WeChat ⁺ `edu_assist_pro` $g_+(\mathbf{x}) - g_-(\mathbf{x})$

Decision boundary for equal covariances $\Lambda_+ = \Lambda_- = \Lambda$ is linear:

Log of ratio of posteriors of two Gaussians: for equal and unequal covariances

$$\begin{aligned}g_-(\mathbf{x}) &= \ln P(C_-) + \frac{1}{2} \ln |\Lambda_-| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_-)^T \Lambda_- (\mathbf{x} - \boldsymbol{\mu}_-) \\&= \ln P(C_-) + \frac{1}{2} \ln |\Lambda_-| - \frac{1}{2} (\mathbf{x}^T \Lambda_- \mathbf{x} - 2\boldsymbol{\mu}_-^T \Lambda_- \mathbf{x} + \boldsymbol{\mu}_-^T \Lambda_- \boldsymbol{\mu}_-) \\g_+(\mathbf{x}) &= \ln P(C_+) + \frac{1}{2} \ln |\Lambda_+| - \frac{1}{2} (\mathbf{x}^T \Lambda_+ \mathbf{x} - 2\boldsymbol{\mu}_+^T \Lambda_+ \mathbf{x} + \boldsymbol{\mu}_+^T \Lambda_+ \boldsymbol{\mu}_+)\end{aligned}$$

For unequal covariance <https://eduassistpro.github.io/>
Your turn: what is the quadratic term in the boundary equation $g_+(\mathbf{x}) - g_-(\mathbf{x})$

Decision boundary for equal covariances $\Lambda_+ = \Lambda_- = \Lambda$ is linear:

$$g_+(\mathbf{x}) - g_-(\mathbf{x}) = \ln \frac{P(C_+)}{P(C_-)} + (\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-)^T \Lambda \mathbf{x} + \mathbf{x}\text{-independent term}$$

Log of ratio of posteriors of two Gaussians: for equal and unequal covariances

$$\begin{aligned}g_-(\mathbf{x}) &= \ln P(C_-) + \frac{1}{2} \ln |\Lambda_-| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_-)^T \Lambda_- (\mathbf{x} - \boldsymbol{\mu}_-) \\&= \ln P(C_-) + \frac{1}{2} \ln |\Lambda_-| - \frac{1}{2} (\mathbf{x}^T \Lambda_- \mathbf{x} - 2\boldsymbol{\mu}_-^T \Lambda_- \mathbf{x} + \boldsymbol{\mu}_-^T \Lambda_- \boldsymbol{\mu}_-) \\g_+(\mathbf{x}) &= \ln P(C_+) + \frac{1}{2} \ln |\Lambda_+| - \frac{1}{2} (\mathbf{x}^T \Lambda_+ \mathbf{x} - 2\boldsymbol{\mu}_+^T \Lambda_+ \mathbf{x} + \boldsymbol{\mu}_+^T \Lambda_+ \boldsymbol{\mu}_+)\end{aligned}$$

For unequal covariance
Your turn: what is the quadratic term in the boundary equation $g_+(\mathbf{x}) - g_-(\mathbf{x})$

Decision boundary for equal covariances $\Lambda_+ = \Lambda_- = \Lambda$ is linear:

$$g_+(\mathbf{x}) - g_-(\mathbf{x}) = \ln \frac{P(C_+)}{P(C_-)} + (\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-)^T \Lambda \mathbf{x} + \mathbf{x}\text{-independent term}$$

Your turn: derive the linear equation

Discriminants (decision boundaries) for 2-class gaussians:
linear if covariances equal, quadratic if not

Assignment Project Exam Help

<https://eduassistpro.github.io/>

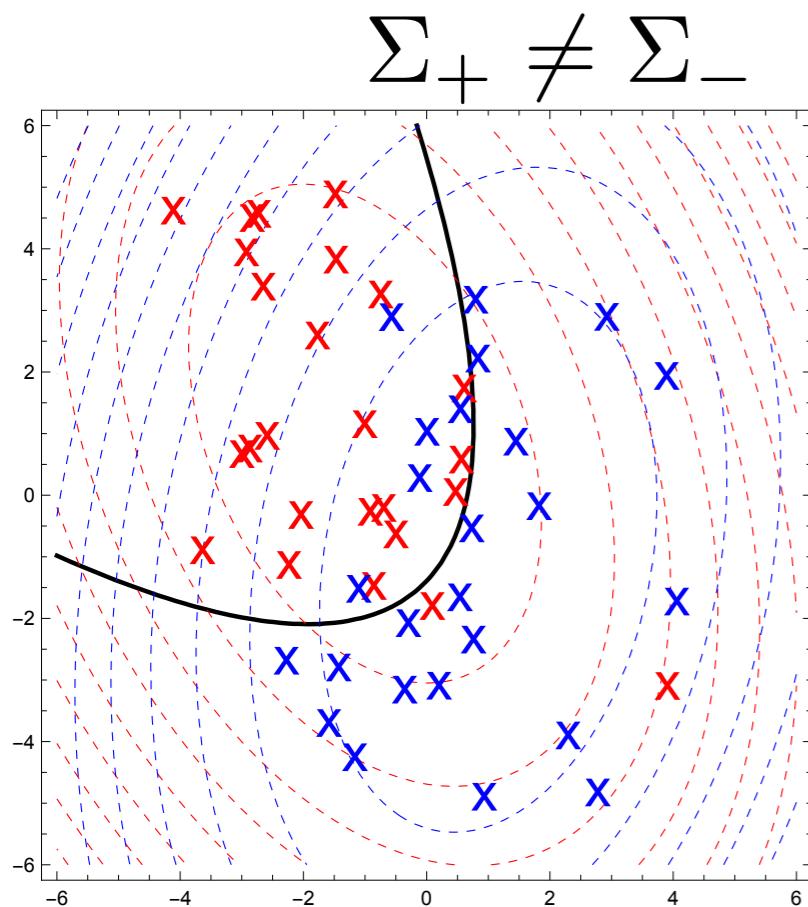
Add WeChat edu_assist_pro

Discriminants (decision boundaries) for 2-class gaussians:
linear if covariances equal, quadratic if not

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

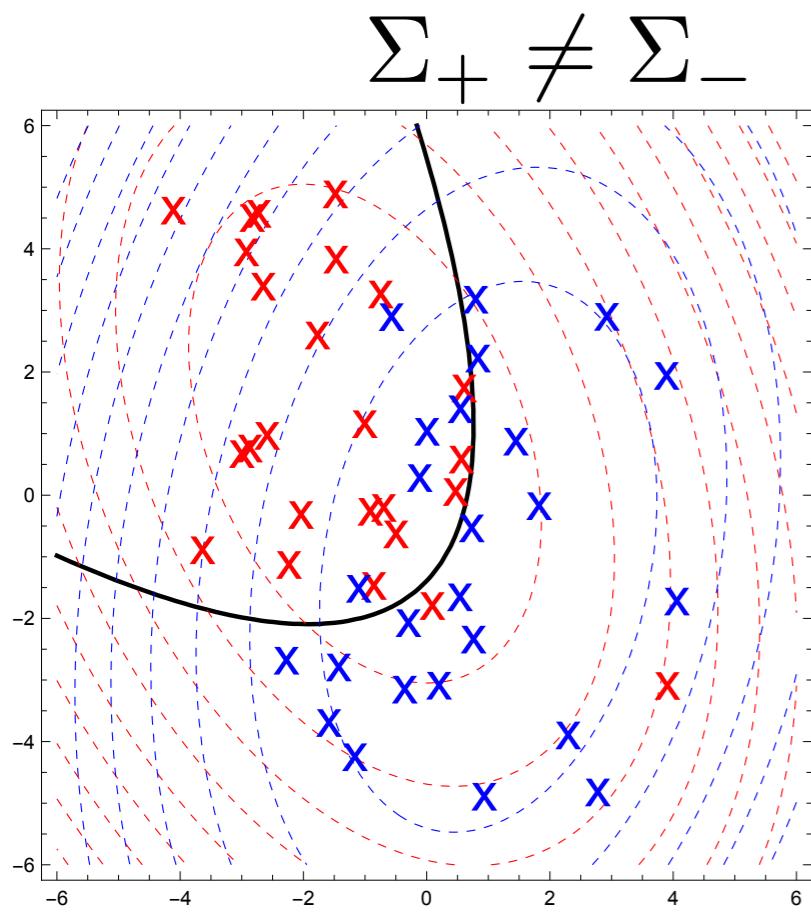


Discriminants (decision boundaries) for 2-class gaussians:
linear if covariances equal, quadratic if not

Assignment Project Exam Help

<https://eduassistpro.github.io/>

If ~~Add WeChat~~ `edu_assist_pro`
matrices are the
same, the
quadratic term is
common to both
classes, and drops
out in the ratio

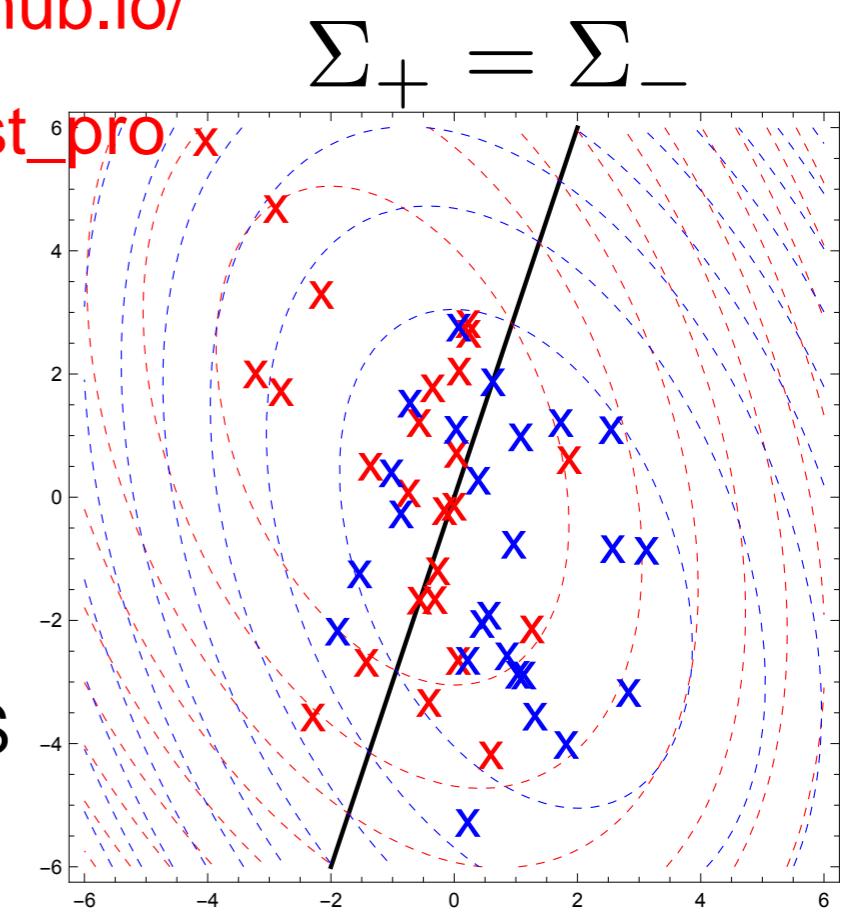
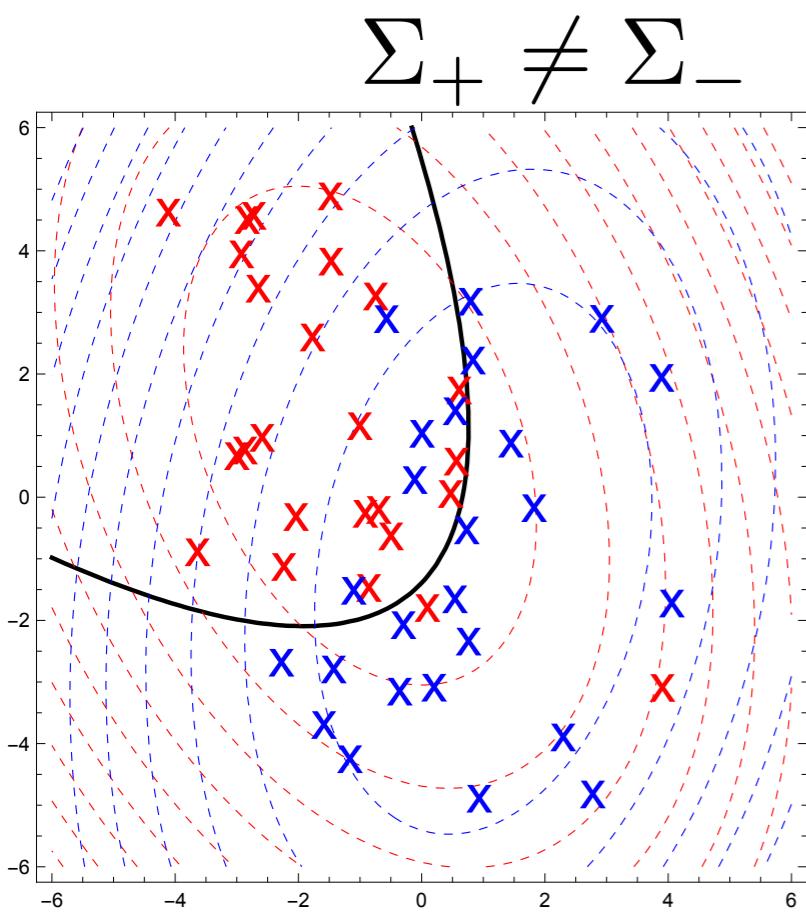


Discriminants (decision boundaries) for 2-class gaussians:
linear if covariances equal, quadratic if not

Assignment Project Exam Help

<https://eduassistpro.github.io/>

If ~~Add WeChat~~ `edu_assist_pro`
the covariance
matrices are the
same, the
quadratic term is
common to both
classes, and drops
out in the ratio



Discriminants for Gaussians for equal priors: data distribution equally divided

1-dim

2-dim

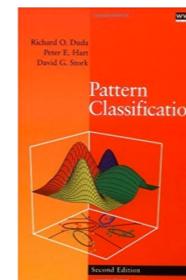
3-dim

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Equal priors: $\log(P_a/P_b) = 0$



From Duda and Hart

Discriminants for Gaussians: unequal priors shifts decision boundary by $\log(P_a/P_b)$

1-dim

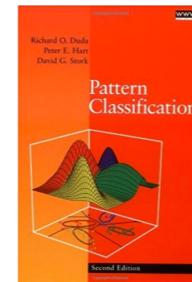
Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

2-dim

3-dim



From Duda and Hart

Discriminants for Gaussians: unequal priors shifts decision boundary by $\log(P_a/P_b)$

1-dim

Prior overrides data distributions

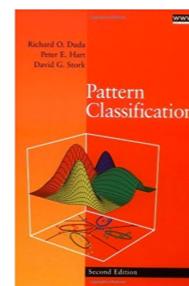
Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

2-dim

3-dim



From Duda and Hart

Discriminants for Gaussians: unequal priors shifts decision boundary by $\log(P_a/P_b)$

1-dim

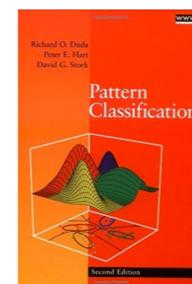
Prior overrides data distributions

Assignment Project Exam Help

<https://eduassistpro.github.io/>
how far the boundary shifts
Add WeChat edu_assist_pro

2-dim

3-dim



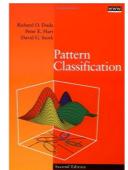
From Duda and Hart

Decision boundaries not perpendicular to line joining means
if covariance matrix not proportional to the identity matrix

$$w = \Sigma^{-1}(\mu_+ - \mu_-)$$

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro



From Duda and Hart

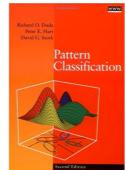
Decision boundaries not perpendicular to line joining means
if covariance matrix not proportional to the identity matrix

$$g_+(\mathbf{x}) - g_-(\mathbf{x}) = \ln \frac{P(C_+)}{P(C_-)} + (\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-)^T \boldsymbol{\Lambda} \mathbf{x} + \mathbf{x}\text{-independent term}$$

Assignment Project Exam Help
 $\omega = \sum \frac{1}{2}(\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-)$

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro



From Duda and Hart

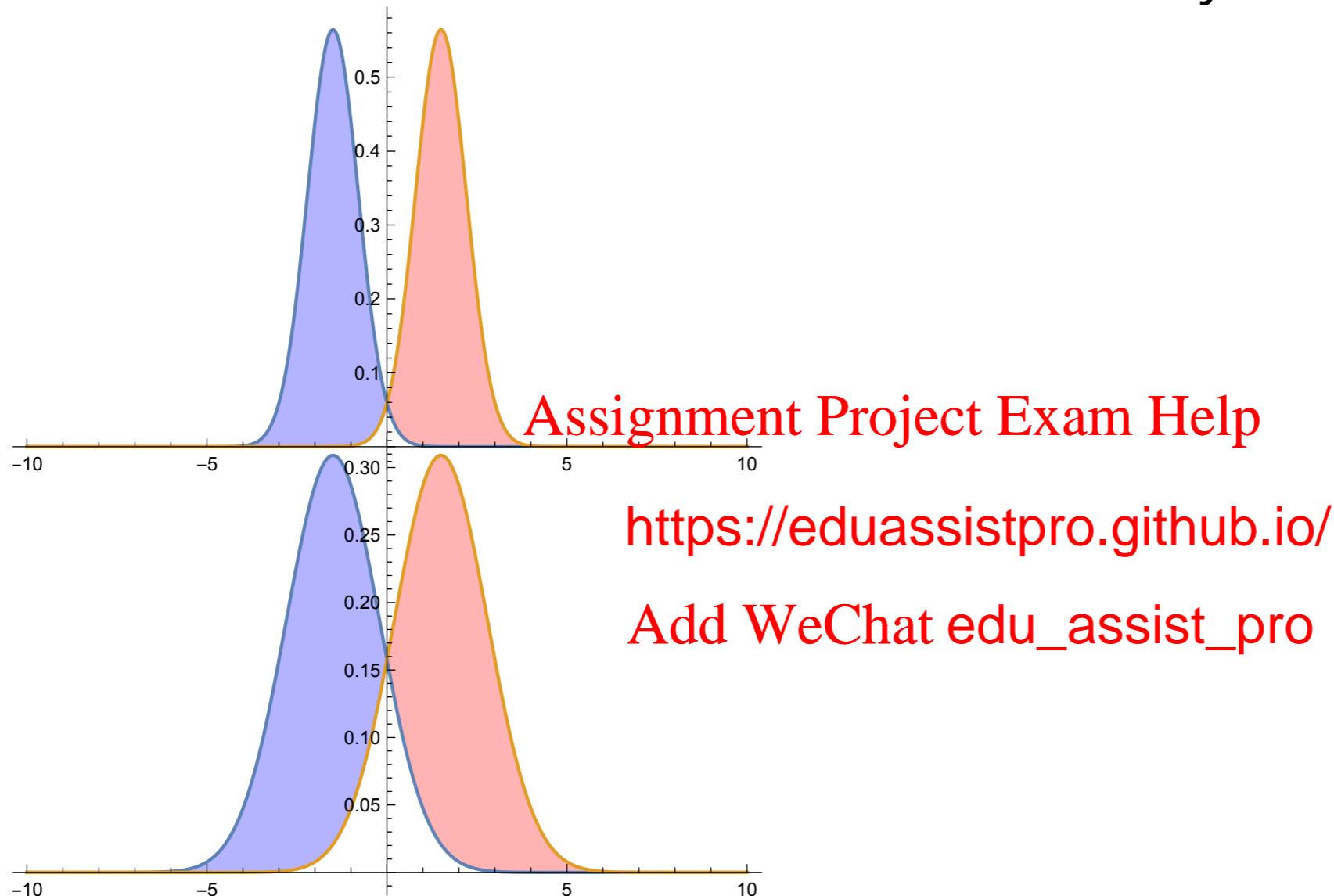
Distinctiveness relative to natural variation measured by distance between class means scaled by standard deviation:

Assignment Project Exam Help

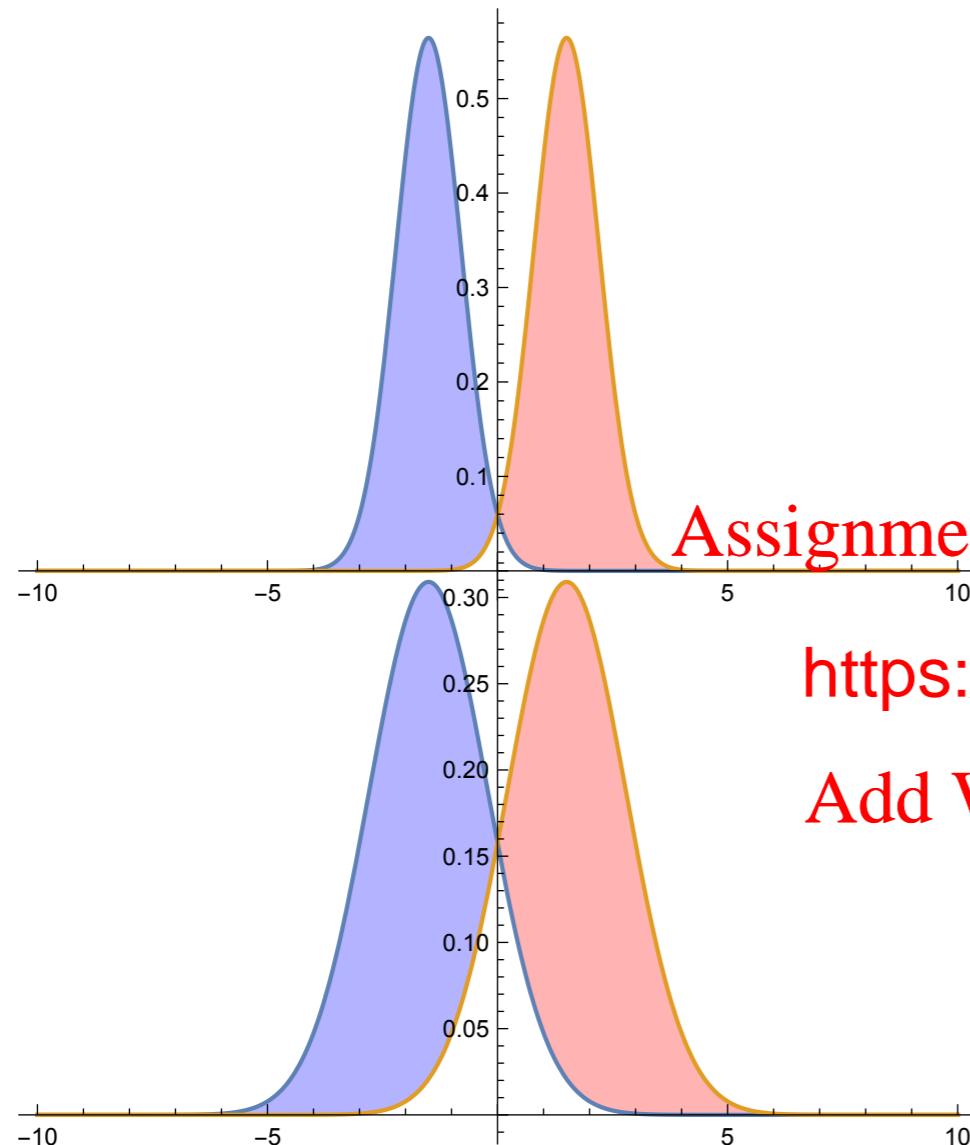
<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Distinctiveness relative to natural variation measured by distance between class means scaled by standard deviation:



Distinctiveness relative to natural variation measured by distance between class means scaled by standard deviation:



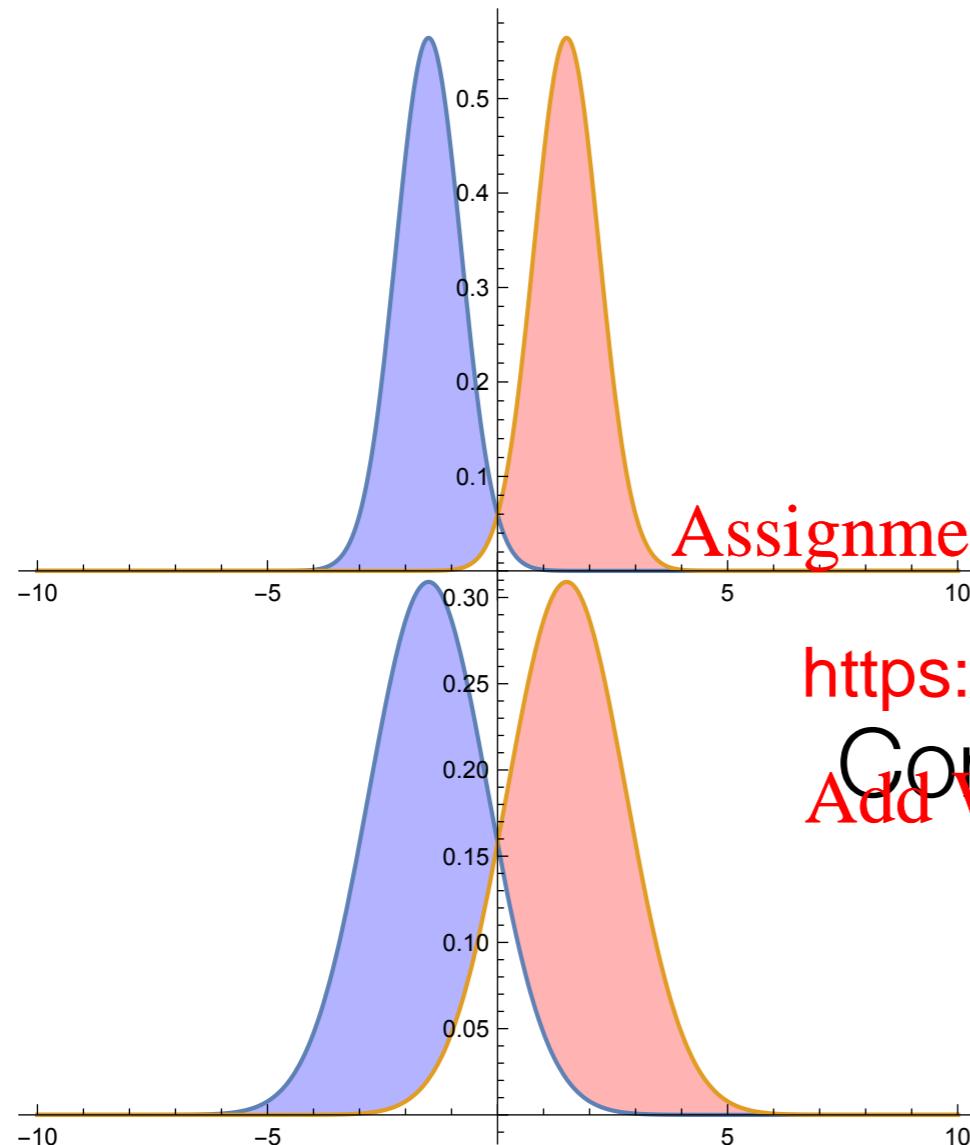
$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Distinctiveness relative to natural variation measured by distance between class means scaled by standard deviation:



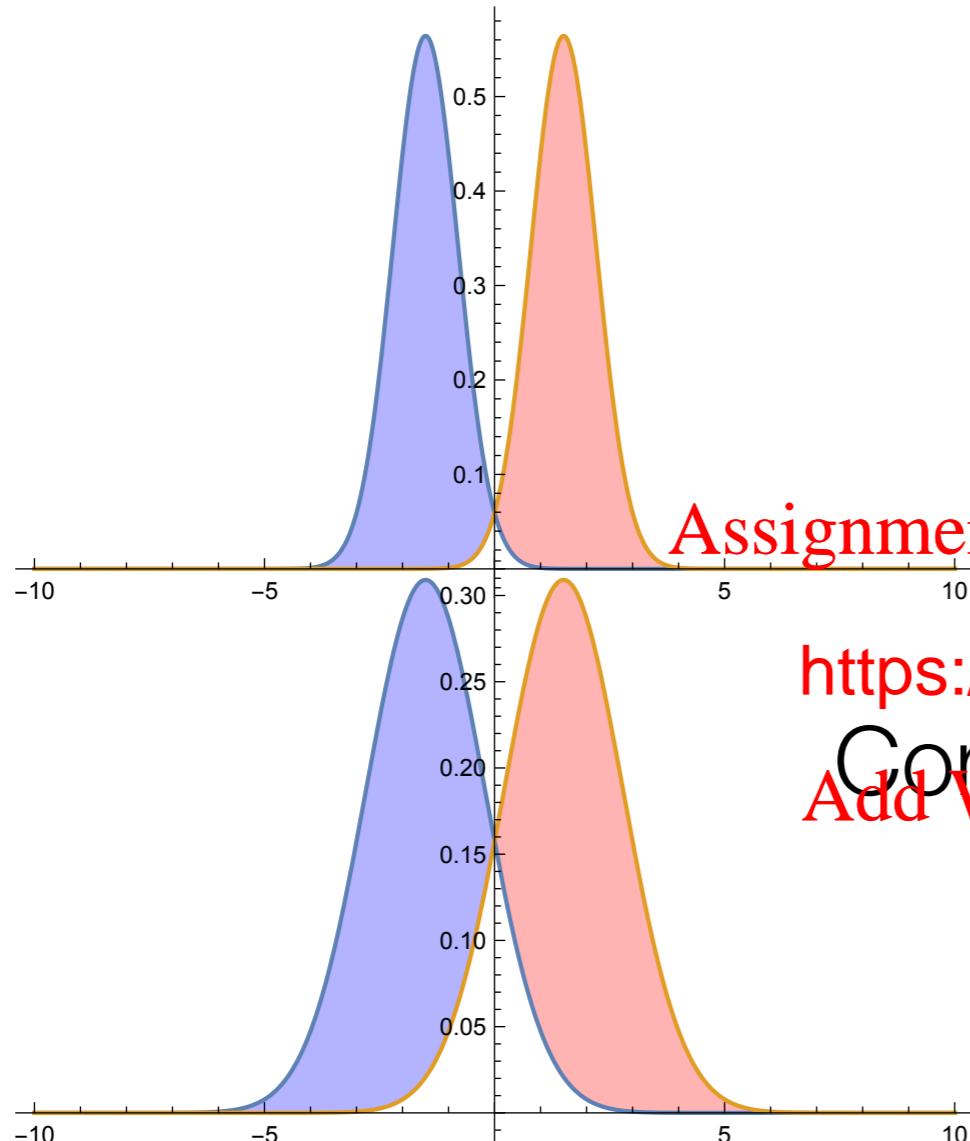
$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Controls lap between classes
Add WeChat edu_assist_pro

Distinctiveness relative to natural variation measured by distance between class means scaled by standard deviation:



$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Controls Add WeChat edu_assist_pro lap between classes

Mahalanobis distance

$$d(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)^T \Sigma^{-1} (\mathbf{x}_1 - \mathbf{x}_2)}$$

Fisher Linear Discriminant Analysis: projections along discriminative directions

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Fisher Linear Discriminant Analysis: projections along discriminative directions

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Fisher Linear Discriminant Analysis: projections along discriminative directions

$$\mu_c = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{x}^n$$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Fisher Linear Discriminant Analysis: projections along discriminative directions

$$\mu_c = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{x}^n \quad m_c = \mathbf{w}^T \mu_c = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{w}^T \mathbf{x}^n$$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Fisher Linear Discriminant Analysis: projections along discriminative directions

$$\mu_c = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{x}^n \quad m_c = \mathbf{w}^T \mu_c = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{w}^T \mathbf{x}^n$$

Assignment Project Exam Help
<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Fisher Linear Discriminant Analysis: projections along discriminative directions

$$\mu_c = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{x}^n \quad m_c = \mathbf{w}^T \mu_c = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{w}^T \mathbf{x}^n$$

Assignment Project Exam Help
<https://eduassistpro.github.io/>

$$S_c = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (y^n - m_c)^2$$

Fisher Linear Discriminant Analysis: projections along discriminative directions

$$\mu_c = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{x}^n \quad m_c = \mathbf{w}^T \mu_c = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{w}^T \mathbf{x}^n$$

Assignment Project Exam Help
<https://eduassistpro.github.io/>

$$S_c = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (y^n - m_c)^2$$

Fisher score:

$$J(\mathbf{w}) = \frac{(m_+ - m_-)^2}{n_+ s_+^2 + n_- s_-^2}$$

Choose $\mathbf{w}^* = \operatorname{argmax}_{\mathbf{w}} J(\mathbf{w})$

Rewrite Fisher score to make weight vector dependence explicit: between and within class scatter

$$s_c^2 = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (y^n - m_c)^2 = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (\mathbf{w}^T \mathbf{x}^n - \mathbf{w}^T \boldsymbol{\mu})^2$$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

denominator:

within class scatter

numerator:

between class scatter

$$J(\mathbf{w}) = \frac{(m_+ - m_-)^2}{n_+ s_+^2 + n_- s_-^2} \implies$$

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}$$

Rewrite Fisher score to make weight vector dependence explicit: between and within class scatter

$$s_c^2 = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (y^n - m_c)^2 = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (\mathbf{w}^T \mathbf{x}^n - \mathbf{w}^T \mu)^2$$

$(\mathbf{a} \cdot \mathbf{b})^2 = (\mathbf{b}^\top \mathbf{a})(\mathbf{a}^\top \mathbf{b})$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

denominator:

within class scatter

numerator:

between class scatter

$$J(\mathbf{w}) = \frac{(m_+ - m_-)^2}{n_+ s_+^2 + n_- s_-^2} \implies$$

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}$$

Rewrite Fisher score to make weight vector dependence explicit: between and within class scatter

$$\begin{aligned}
 s_c^2 &= \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (y^n - m_c)^2 = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (\mathbf{w}^T \mathbf{x}^n - \mathbf{w}^T \mu)^2 \\
 &= \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{w}^T (\mathbf{x}^n - \mu) (\mathbf{x}^n - \mu)^T \mathbf{w}
 \end{aligned}$$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

denominator:

within class scatter

numerator:

between class scatter

$$J(\mathbf{w}) = \frac{(m_+ - m_-)^2}{n_+ s_+^2 + n_- s_-^2} \implies$$

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}$$

Rewrite Fisher score to make weight vector dependence explicit: between and within class scatter

$$\begin{aligned}
 s_c^2 &= \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (y^n - m_c)^2 = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (\mathbf{w}^T \mathbf{x}^n - \mathbf{w}^T \mu)^2 \\
 &= \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{w}^T (\mathbf{x}^n - \mu) (\mathbf{x}^n - \mu)^T \mathbf{w} \\
 &\triangleq \mathbf{w}^T \mathbf{S} \mathbf{w}
 \end{aligned}$$

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

denominator:

within class scatter

numerator:

between class scatter

$$J(\mathbf{w}) = \frac{(m_+ - m_-)^2}{n_+ s_+^2 + n_- s_-^2} \implies$$

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}$$

Rewrite Fisher score to make weight vector dependence explicit: between and within class scatter

$$\begin{aligned}
 s_c^2 &= \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (y^n - m_c)^2 = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (\mathbf{w}^T \mathbf{x}^n - \mathbf{w}^T \mu)^2 \\
 &= \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{w}^T (\mathbf{x}^n - \mu) (\mathbf{x}^n - \mu)^T \mathbf{w} \\
 &\triangleq \mathbf{w}^T \mathbf{S} \mathbf{w} \quad \text{Assignment Project Exam Help}
 \end{aligned}$$

$$\begin{aligned}
 n_+ s_+^2 + n_- s_-^2 &= \mathbf{w}^T \mathbf{S}_W \mathbf{w} \\
 &\triangleq \mathbf{w}^T \mathbf{S}_W \mathbf{w}
 \end{aligned}$$

denominator:
within class scatter

numerator:
between class scatter

$$J(\mathbf{w}) = \frac{(m_+ - m_-)^2}{n_+ s_+^2 + n_- s_-^2} \implies$$

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}$$

Rewrite Fisher score to make weight vector dependence explicit: between and within class scatter

$$\begin{aligned}
 s_c^2 &= \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (y^n - m_c)^2 = \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} (\mathbf{w}^T \mathbf{x}^n - \mathbf{w}^T \mu)^2 \\
 &= \frac{1}{n_c} \sum_{\mathbf{x}^n \in \mathcal{D}_c} \mathbf{w}^T (\mathbf{x}^n - \mu) (\mathbf{x}^n - \mu)^T \mathbf{w} \\
 &\triangleq \mathbf{w}^T \mathbf{S} \mathbf{w} \quad \text{Assignment Project Exam Help}
 \end{aligned}$$

$$\begin{aligned}
 n_+ s_+^2 + n_- s_-^2 &= \textcolor{red}{\frac{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}} \quad \text{denominator:} \\
 &\triangleq \textcolor{red}{\frac{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}} \quad \text{within class scatter}
 \end{aligned}$$

$$\begin{aligned}
 (m_+ - m_-)^2 &= (\mathbf{w}^T (\mu_+ - \mu_-))^2 \\
 &= \mathbf{w}^T (\mu_+ - \mu_-) (\mu_+ - \mu_-)^T \mathbf{w} \quad \text{numerator:} \\
 &\quad \text{between class scatter}
 \end{aligned}$$

$$J(\mathbf{w}) = \frac{(m_+ - m_-)^2}{n_+ s_+^2 + n_- s_-^2} \implies$$

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}$$

Optimisation: find best direction, set gradients to 0

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Optimisation: find best direction, set gradients to 0

Partial derivatives (gradient) set to 0

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \frac{\text{Assignment Project Exam Help}}{(\mathbf{w}^T \mathbf{S}_B \mathbf{w})} - (\mathbf{w}^T \mathbf{S}_B \mathbf{w}) \mathbf{S}_W \mathbf{w}$$

reduces to a generalised eigenvalue problem.

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w}$$

Optimisation: find best direction, set gradients to 0

$$\left(\frac{u}{v}\right)' = \frac{1}{v^2} (u'v - uv')$$

Partial derivatives (gradient) set to 0

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \frac{\text{Assignment Project Exam Help}}{(\mathbf{w}^T \mathbf{S}_B \mathbf{w})} (\mathbf{S}_B \mathbf{w} - (\mathbf{w}^T \mathbf{S}_B \mathbf{w}) \mathbf{S}_W \mathbf{w})$$

reduces to a generalised eigenvalue problem

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w}$$

Optimisation: find best direction, set gradients to 0

$$\left(\frac{u}{v}\right)' = \frac{1}{v^2} (u'v - uv')$$

Partial derivatives (gradient) set to 0

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \frac{\text{Assignment Project Exam Help}}{(\mathbf{w}^T \mathbf{S}_B \mathbf{w})} (\mathbf{S}_B \mathbf{w} - \underline{(\mathbf{w}^T \mathbf{S}_B \mathbf{w}) \mathbf{S}_W \mathbf{w}})$$

reduces to a **generalised eigenvalue**

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w}$$

Optimisation: find best direction, set gradients to 0

Partial derivatives (gradient) set to 0

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \frac{\text{Assignment Project Exam Help}}{\text{https://eduassistpro.github.io/}} \left(\mathbf{w}^T \mathbf{S}_B \mathbf{w} - (\mathbf{w}^T \mathbf{S}_B \mathbf{w}) \mathbf{S}_W \mathbf{w} \right)$$

reduces to a generalised eigenvalue

$$\mathbf{S}_B \mathbf{w} = \underline{\lambda} \mathbf{S}_W \mathbf{w}$$

$$\left(\frac{u}{v} \right)' = \frac{1}{v^2} (u'v - uv')$$

scalars

ratio

Optimisation: find best direction, set gradients to 0

Partial derivatives (gradient) set to 0

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \frac{\text{Assignment Project Exam Help}}{\text{https://eduassistpro.github.io/}} \left(\mathbf{w}^T \mathbf{S}_B \mathbf{w} - (\mathbf{w}^T \mathbf{S}_B \mathbf{w}) \mathbf{S}_W \mathbf{w} \right)$$

reduces to a generalised eigenvalue

$$\mathbf{S}_B \mathbf{w} = \underline{\lambda} \mathbf{S}_W \mathbf{w}$$

ratio

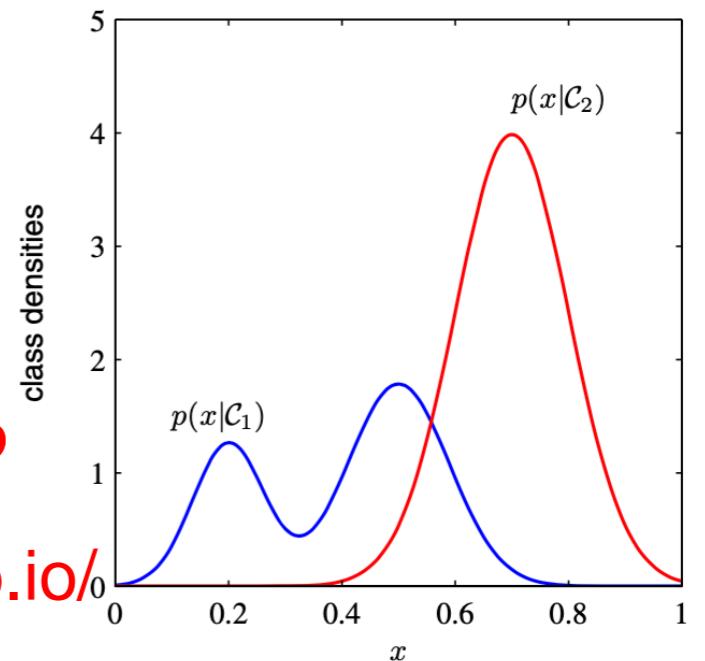
Extend to multiple classes, not just 2.

Classifiers can be generative or discriminative

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro



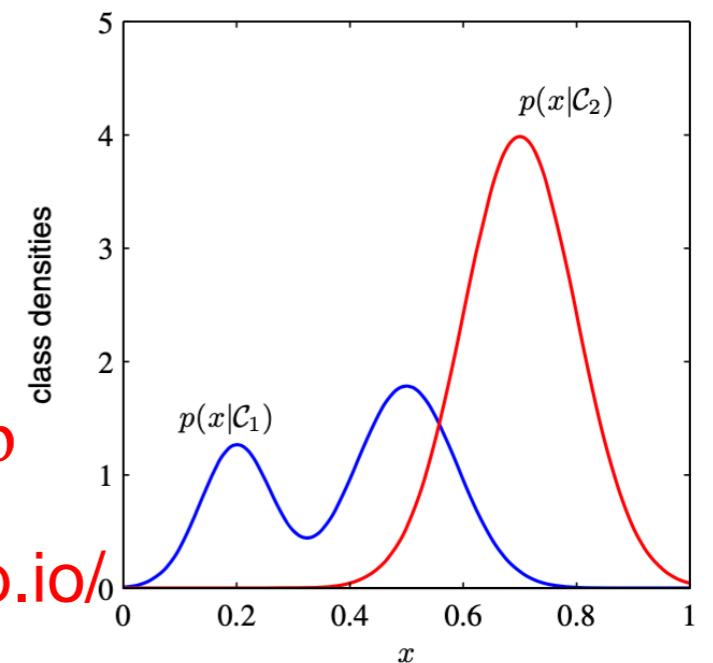
Classifiers can be generative or discriminative

- Generative – class conditional probabilities can be used to simulate (generate) data

Assignment Project Exam Help

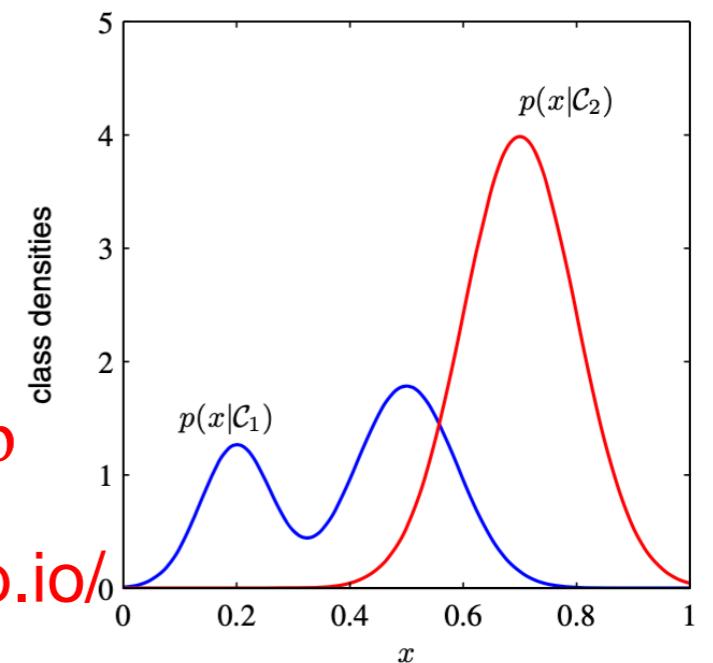
<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro



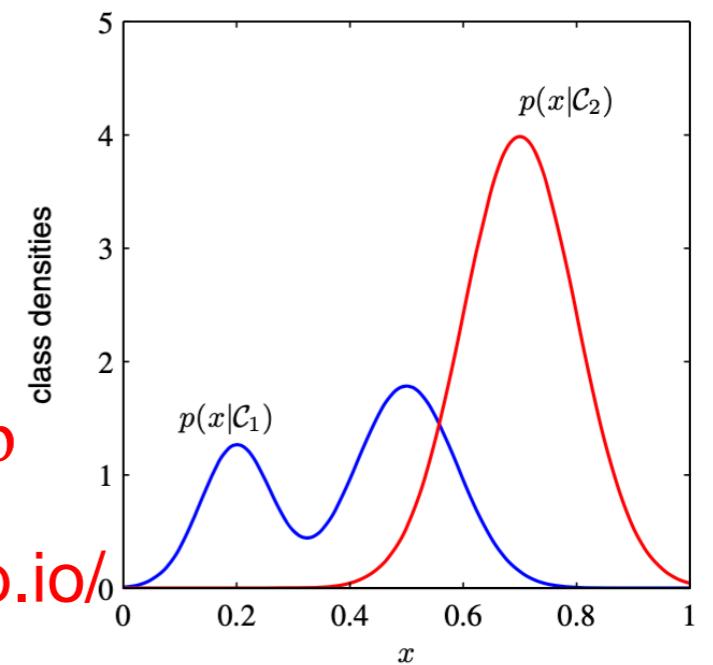
Classifiers can be generative or discriminative

- Generative – class conditional probabilities can be used to simulate (generate) data
 - Generative – probabilities consistent with allocation from decision boundaries
- Assignment Project Exam Help
<https://eduassistpro.github.io/>
Add WeChat edu_assist_pro



Classifiers can be generative or discriminative

- Generative – class conditional probabilities can be used to simulate (generate) data
 - Generative – probabilities consistent with allocation from decision boundaries
 - Discriminative – does not model data within classes, only what distinguishes a class from another
- Assignment Project Exam Help
<https://eduassistpro.github.io/>
Add WeChat edu_assist_pro



Classifiers can be generative or discriminative

- Generative – class conditional probabilities can be used to simulate (generate) data

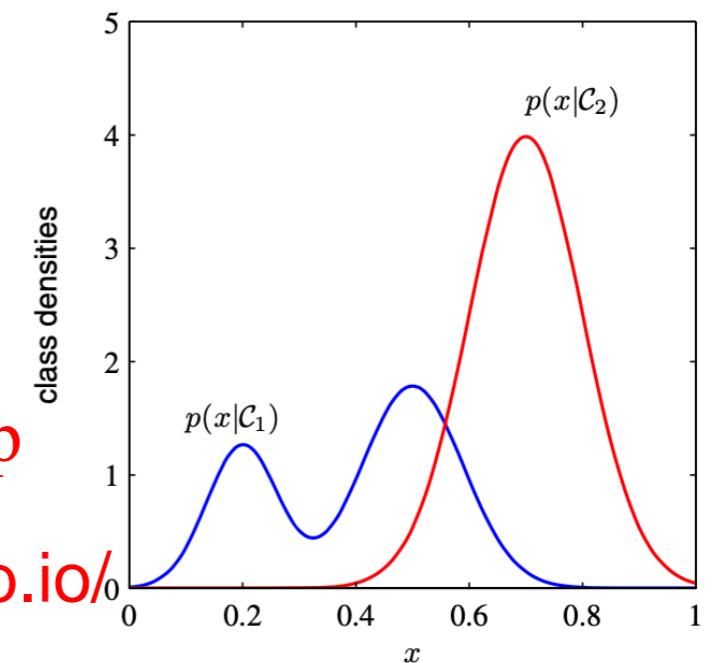
Assignment Project Exam Help

- Generative – probabilities consistent with allocation <https://eduassistpro.github.io/> from decision boundaries

Add WeChat edu_assist_pro

- Discriminative – does not model data within classes, only what distinguishes a class from another

- Next time: maximum margin classifiers (Support Vector Machines)



References

- PRML, Bishop, Chapter 1, section 1.5
- PRML, Bishop, Chapter 4.
[Assignment](#) [Project](#) [Exam](#) [Help](#)
- FCML, Rogers-Gir <https://eduassistpro.github.io/>
[Add WeChat](#) [edu_assist_pro](#)