German University in Cairo
Faculty of Media Engineering and Technology
Winter 2019

CSEN 1022 – Machine Learning

Problem Set #5 Solution

Problem 1

Consider the data given below about the weight and height of different individuals:

ID	Weight (Kg)	Height (cm)	Gender?
1	65	155	Female
2	60	157	Female
3	63	153	Female
4	72	159	Female
5	75	165	Male
6	70	162	Male
7	80	180	Male
8	75	173	Male
9	85	170	Male

Using **Naïve Bayes classifier**, predict the gender of an individual with weight of 75 Kg and height 170 cm. Note that since the data is continuous, you can use the Gaussian distribution $N(x|\mu,\sigma)$ for each feature that takes the form

$$N(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \text{ where } \mu = \frac{1}{M}\sum_{i=1}^{M} x_i \text{ and } \sigma^2 = \frac{1}{M}\sum_{i=1}^{M}(x_i-\mu)^2$$

Solution:

The point we want to predict is $\mathbf{x}$ where $\mathbf{x} = (75, 170)$ given that the first dimension/feature is the weight and the second dimension/feature is the height.

We first define two classes C_1 and C_2, corresponding to Gender = Female and Gender = Male, respectively. To classify the gender of the point $\mathbf{x}$, we need to compute $p(C_1|\mathbf{x}) = p(\text{Gender} = \text{Female}|\mathbf{x})$ and $p(C_2|\mathbf{x}) = p(\text{Gender} = \text{Male}|\mathbf{x})$.

and find which conditional probability is larger. If the first one is larger, then our prediction is Gender = Female. If the second one is larger, then our prediction is Gender = Male.

Since $p(C_1|\mathbf{x}) = \frac{p(\mathbf{x}|C_1)p(C_1)}{p(\mathbf{x})}$

We need to compute $p(\mathbf{x}|C_1) = p(\mathbf{x}|\text{ Gender} = \text{Female})$. Using the Naïve Bayes assumption which assumes that the dimensions of the input data (the weight and height) are independent, we can re-write $p(\mathbf{x}|C_1)$ as $p(\mathbf{x}|C_1) = \prod_{i=1}^{D} p(x_i|C_1)$

$p(\mathbf{x}|\text{ Gender} = \text{Female}) = p(\text{weight} = 75 \mid \text{Gender} = \text{Female})\, p(\text{Height} = 170 \mid \text{Gender} = \text{Female})$

Similarly, $p(\mathbf{x}|C_2)$ can be re-written as $p(\mathbf{x}|C_2) = \prod_{i=1}^{D} p(x_i|C_2)$

$p(\mathbf{x}|\text{ Gender} = \text{Male}) = p(\text{weight} = 75 \mid \text{Gender} = \text{Male})\, p(\text{Height} = 170 \mid \text{Gender} = \text{Male})$

CSEN 1022 – Machine Learning

Problem Set #5 Solution

Since we have the Gaussian distribution $N(x|\mu,\sigma)$ for each feature that takes the form

$$N(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \text{ where } \mu = \frac{1}{M}\sum_{i=1}^{M} x_i \text{ and } \sigma^2 = \frac{1}{M}\sum_{i=1}^{M}(x_i-\mu)^2$$

Using the above parameters, we need to calculate μ and σ^2 for each feature in each class in order to calculate $p(\mathbf{x} | C_1)$ and $p(\mathbf{x} | C_2)$.

The required calculations are shown below:

$$\mu_{weight,female} = \frac{1}{M_{female}} \sum_{i=1}^{M_{female}} x_i = \frac{1}{4}(65+60+63+72) = 65$$

$$\mu_{height,female} = \frac{1}{M_{female}} \sum_{i=1}^{M_{female}} x_i = \frac{1}{4}(155+157+153+159) = 156$$

$$\mu_{weight,male} = \frac{1}{M_{male}} \sum_{i=1}^{M_{male}} x_i = \frac{1}{5}(75+70+80+75+85) = 77$$

$$\mu_{height,male} = \frac{1}{M_{male}} \sum_{i=1}^{M_{male}} x_i = \frac{1}{5}(165+162+180+173+170) = 170$$

$$\sigma^2_{weight,female} = \frac{1}{M_{female}} \sum_{i=1}^{M_{female}} \left(x_i - \mu_{weight,female}\right)^2$$
$$= \frac{1}{4}[(65-65)^2 + (60-65)^2 + (63-65)^2 + (72-65)^2] = 19.5$$

$$\sigma^2_{height,female} = \frac{1}{M_{female}} \sum_{i=1}^{M_{female}} \left(x_i - \mu_{height,female}\right)^2$$
$$= \frac{1}{4}[(155-156)^2 + (157-156)^2 + (153-156)^2 + (159-156)^2] = 5$$

$$\sigma^2_{weight,male} = \frac{1}{M_{male}} \sum_{i=1}^{M_{male}} \left(x_i - \mu_{weight,male}\right)^2$$
$$= \frac{1}{5}[(75-77)^2 + (70-77)^2 + (80-77)^2 + (75-77)^2 + (85-77)^2] = 26$$

$$\sigma^2_{height,male} = \frac{1}{M_{male}} \sum_{i=1}^{M_{male}} \left(x_i - \mu_{height,male}\right)^2$$
$$= \frac{1}{5}[(165-170)^2 + (162-170)^2 + (180-170)^2 + (173-170)^2 + (170-170)^2] = 39.6$$

Therefore, the probabilities can be calculated as follows:

$$p(\text{weight} \ = \ 75 \,|\, \text{Gender} \ = \ \text{Female}) \ = \frac{1}{\sqrt{2\pi\sigma^2_{weight,female}}} exp\left(-\frac{\left(x-\mu_{weight,female}\right)^2}{2\sigma^2_{weight,female}}\right)$$
$$= \frac{1}{\sqrt{2\pi * 19.5}} exp\left(-\frac{(75-65)^2}{2 * 19.5}\right) = 0.0070$$

$$p(\text{height} \ = \ 170 \,|\, \text{Gender} \ = \ \text{Female}) \ = \frac{1}{\sqrt{2\pi\sigma^2_{height,female}}} exp\left(-\frac{\left(x-\mu_{height,female}\right)^2}{2\sigma^2_{height,female}}\right)$$
$$= \frac{1}{\sqrt{2\pi * 5}} exp\left(-\frac{(170-156)^2}{2 * 5}\right) = 5.486 * 10^{-10}$$

$$p(\text{weight} \ = \ 75 \,|\, \text{Gender} \ = \ \text{Male}) \ = \frac{1}{\sqrt{2\pi\sigma^2_{weight,male}}} exp\left(-\frac{\left(x-\mu_{weight,male}\right)^2}{2\sigma^2_{weight,male}}\right)$$
$$= \frac{1}{\sqrt{2\pi * 26}} exp\left(-\frac{(75-77)^2}{2 * 26}\right) = 0.0724$$

$$p(\text{height} \ = \ 170 \,|\, \text{Gender} \ = \ \text{Male}) \ = \frac{1}{\sqrt{2\pi\sigma^2_{height,male}}} exp\left(-\frac{\left(x-\mu_{height,male}\right)^2}{2\sigma^2_{height,male}}\right)$$
$$= \frac{1}{\sqrt{2\pi * 39.6}} exp\left(-\frac{(170-170)^2}{2 * 39.6}\right) = 0.0634$$

$$p(\text{Gender} \ = \ \text{Female}) = \frac{4}{9}$$

$$p(\text{Gender} \ = \ \text{Male}) = \frac{5}{9}$$

Problem Set #5 Solution

Therefore,

$p(\mathbf{x}|$ Gender = Female$) = 0.0070 * 5.486 * 10^{-10} = 3.8402 * 10^{-12}$

$p(\mathbf{x}|$ Gender = Male$) = 0.0724 * 0.0634 = 0.0046$

$$\frac{p(C_1|\mathbf{x})}{p(C_2|\mathbf{x})} = \frac{p(\mathbf{x}|Gender = Female)p(Gender = Female)}{p(\mathbf{x}|Gender = Male)p(Gender = Male)} = \frac{1.7068 * 10^{-12}}{0.0026} = 6.564 * 10^{-10}$$

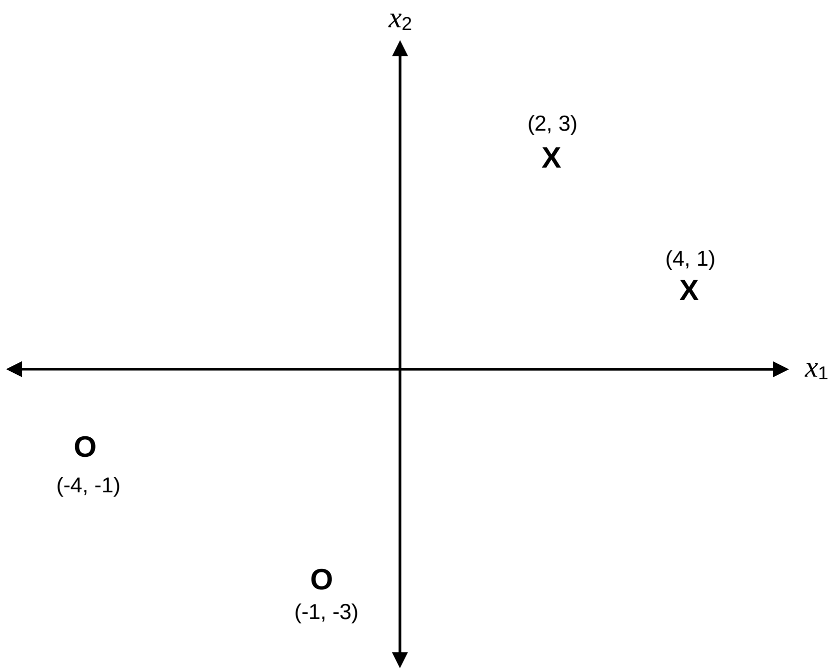
Since $\frac{p(C_1|\mathbf{x})}{p(C_2|\mathbf{x})} < 1$ then $\mathbf{x} \in C_2$. Therefore, the gender of testing point $\mathbf{x}$ is male.

CSEN 1022 – Machine Learning

Problem Set #5 Solution

Problem 2

For the data given below, use the maximum likelihood estimate of Gaussian generative model to find the decision boundary.



Solution:

- Let the target value for class C_1 (**X**) be $t = 1$ and for class C_2 (**O**) be $t = 0$.
- The prior probability of class C_1 is

$$p(C_1) = \pi = \frac{1}{4}\sum_{n=1}^{4} t_n = \frac{1}{4}(1 + 1 + 0 + 0) = 0.5$$

- The prior probability of class C_2 is thus

$$p(C_2) = 1 - \pi = 0.5$$

- The mean of the input vectors in class C_1 is

$$\boldsymbol{\mu}_1 = \frac{1}{N_1}\sum_{n=1}^{4} t_n \boldsymbol{x}_n = \frac{1}{2}\left(1 \times \begin{bmatrix}2\\3\end{bmatrix} + 1 \times \begin{bmatrix}4\\1\end{bmatrix} + 0 \times \begin{bmatrix}-1\\-3\end{bmatrix} + 0 \times \begin{bmatrix}-4\\-1\end{bmatrix}\right) = \begin{bmatrix}3\\2\end{bmatrix}$$

- The mean of the input vectors in class C_2 is

$$\boldsymbol{\mu}_2 = \frac{1}{N_2}\sum_{n=1}^{4} (1 - t_n) \boldsymbol{x}_n = \frac{1}{2}\left(0 \times \begin{bmatrix}2\\3\end{bmatrix} + 0 \times \begin{bmatrix}4\\1\end{bmatrix} + 1 \times \begin{bmatrix}-1\\-3\end{bmatrix} + 1 \times \begin{bmatrix}-4\\-1\end{bmatrix}\right) = \begin{bmatrix}-2.5\\-2\end{bmatrix}$$

CSEN 1022 – Machine Learning

Problem Set #5 Solution

- The covariance matrix Σ is given by

$$\Sigma = \frac{N_1}{N}\boldsymbol{S}_1 + \frac{N_2}{N}\boldsymbol{S}_2 = 0.5\boldsymbol{S}_1 + 0.5\boldsymbol{S}_2$$

$$\boldsymbol{S}_1 = \sum_{\boldsymbol{x}_n \in C_1} (\boldsymbol{x}_n - \boldsymbol{\mu}_1)(\boldsymbol{x}_n - \boldsymbol{\mu}_1)^T$$

$$= \left(\begin{bmatrix}2\\3\end{bmatrix} - \begin{bmatrix}3\\2\end{bmatrix}\right)\left(\begin{bmatrix}2\\3\end{bmatrix} - \begin{bmatrix}3\\2\end{bmatrix}\right)^T + \left(\begin{bmatrix}4\\1\end{bmatrix} - \begin{bmatrix}3\\2\end{bmatrix}\right)\left(\begin{bmatrix}4\\1\end{bmatrix} - \begin{bmatrix}3\\2\end{bmatrix}\right)^T$$

$$= \begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} + \begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} = \begin{bmatrix}2 & -2\\-2 & 2\end{bmatrix}$$

$$\boldsymbol{S}_2 = \sum_{\boldsymbol{x}_n \in C_2} (\boldsymbol{x}_n - \boldsymbol{\mu}_2)(\boldsymbol{x}_n - \boldsymbol{\mu}_2)^T$$

$$= \left(\begin{bmatrix}-1\\-3\end{bmatrix} - \begin{bmatrix}-2.5\\-2\end{bmatrix}\right)\left(\begin{bmatrix}-1\\-3\end{bmatrix} - \begin{bmatrix}-2.5\\-2\end{bmatrix}\right)^T + \left(\begin{bmatrix}-4\\-1\end{bmatrix} - \begin{bmatrix}-2.5\\-2\end{bmatrix}\right)\left(\begin{bmatrix}-4\\-1\end{bmatrix} - \begin{bmatrix}-2.5\\-2\end{bmatrix}\right)^T$$

$$= \begin{bmatrix}2.25 & -1.5\\-1.5 & 1\end{bmatrix} + \begin{bmatrix}2.25 & -1.5\\-1.5 & 1\end{bmatrix} = \begin{bmatrix}4.5 & -3\\-3 & 2\end{bmatrix}$$

$$\Sigma = 0.5\boldsymbol{S}_1 + 0.5\boldsymbol{S}_2 = \begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} + \begin{bmatrix}2.25 & -1.5\\-1.5 & 1\end{bmatrix} = \begin{bmatrix}3.25 & -2.5\\-2.5 & 2\end{bmatrix}$$

- Given that the equation of the decision boundary is $\mathbf{w}^T\mathbf{x} + w_0 = 0$ where

$$\boldsymbol{w} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

$$w_0 = -\frac{1}{2}\boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^T\Sigma^{-1}\boldsymbol{\mu}_2 + ln\frac{p(C_1)}{p(C_2)}$$

Therefore

$$\boldsymbol{w} = \begin{bmatrix}3.25 & -2.5\\-2.5 & 2\end{bmatrix}^{-1}\left(\begin{bmatrix}3\\2\end{bmatrix} - \begin{bmatrix}-2.5\\-2\end{bmatrix}\right) = \begin{bmatrix}8 & 10\\10 & 13\end{bmatrix}\begin{bmatrix}5.5\\4\end{bmatrix} = \begin{bmatrix}84\\107\end{bmatrix}$$

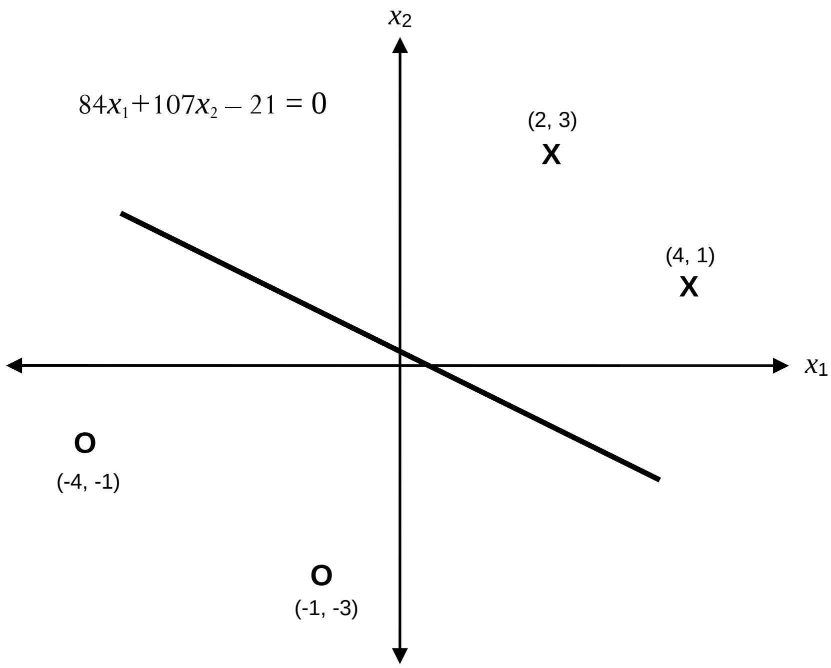
$$w_0 = -0.5[3 \quad 2]\begin{bmatrix}8 & 10\\10 & 13\end{bmatrix}\begin{bmatrix}3\\2\end{bmatrix} + 0.5[-2.5 \quad -2]\begin{bmatrix}8 & 10\\10 & 13\end{bmatrix}\begin{bmatrix}-2.5\\-2\end{bmatrix} + ln\frac{0.5}{0.5}$$

$$= -21$$

Problem Set #5 Solution

The equation of the decision boundary is thus given by

$\mathbf{w}^T\mathbf{x} + w_0 = 0 \rightarrow 84x_1 + 107x_2 - 21 = 0$

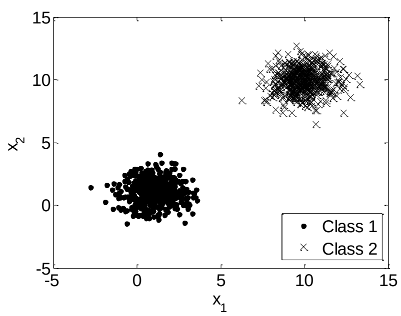
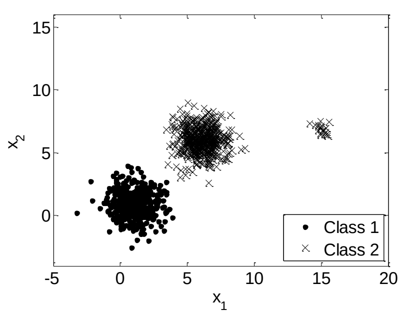
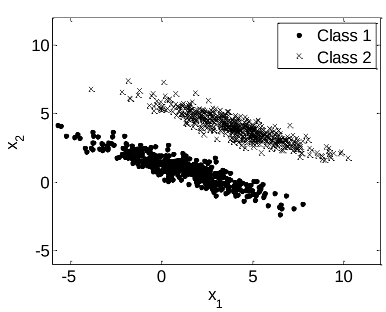


CSEN 1022 – Machine Learning

Problem Set #5 Solution

Problem 3

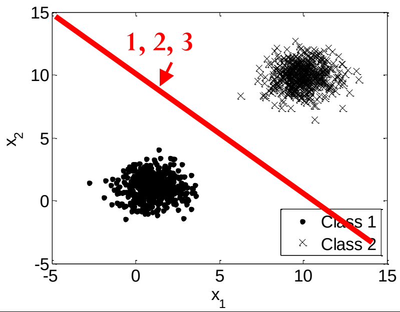
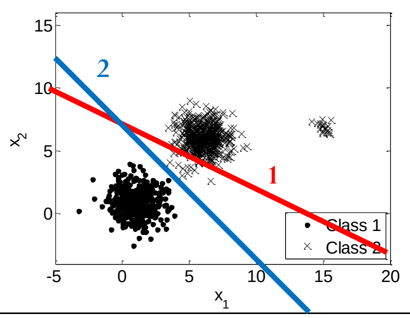
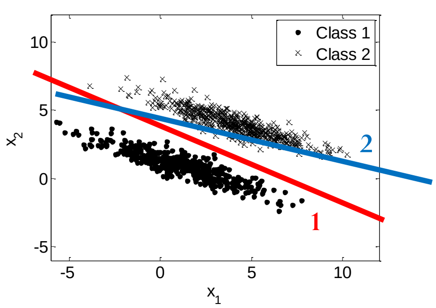
For each of the datasets shown in the left column, approximately **sketch** the decision boundary corresponding to each of the classifiers listed in the right column and briefly **state** below why they perform this way. You can sketch your decision boundaries on the same figures in the left column.

<table>
<tr><td>A)

</td><td>1. Least Squares Classifier
2. Naïve Bayes Classifier
3. Gaussian Generative Model</td></tr>
<tr><td>B)

</td><td>1. Least Squares Classifier
2. Perceptron</td></tr>
<tr><td>C)

</td><td>1. Least Squares Classifier
2. Naïve Bayes Classifier</td></tr>
</table>

CSEN 1022 – Machine Learning

Problem Set #5 Solution

Solution:

A) 	1. Least Squares Classifier 2. Naïve Bayes Classifier 3. Gaussian Generative Model
B) 	1. Least Squares Classifier 2. Perceptron
C) 	1. Least Squares Classifier 2. Naïve Bayes Classifier

A) All classifiers behave correctly:
- Least Squares gives such decision boundary since the data is clean with 2 isolated classes and no outliers.
- Naïve Bayes works well since the 2 dimensions are independent.
- Gaussian Generative Model works well given the Gaussian distribution of the two classes.

B) Least Squares wouldn't be able to classify the 2 classes with no miss-classification since it is sensitive to outliers as it penalizes points that are too correct.
Perceptron performs well as shown since it is robust to outliers.

Problem Set #5 Solution

C) Least Squares performs well same as in (A) (no outliers)
Naïve Bayes miss-classifies some points since the dimensions are not independent (the 2 dimensions are negatively correlated). The covariance matrix in this case is not diagonal.