SOLUTIONS

King's College London

This paper is part of an examination of the College counting towards the award of a degree. Examinations are governed by the College Regulations under the authority of the Academic Board.

Degree Programmes	MSc, MSci
Module Code	7CCSMPNN
Module Title	Pattern Recognition
Examination Period	May 2016 (Period 2)
Time Allowed	Three hours
Rubric	ANSWER QUESTION ONE AND ANY THREE OUT OF THE REMAINING FIVE QUESTIONS. ANSWER EACH QUESTION ON A NEW PAGE OF YOUR ANSWER BOOK AND WRITE ITS NUMBER IN THE SPACE PROVIDED. All questions carry equal marks. If more than three questions other than Question One are answered, clearly indicate which answers you would like to be marked on the front page of the answer booklet, otherwise, Question One and the first three subsequent questions answered, in exam paper order, will be marked.
Calculators	Calculators may be used. The following models are permitted: Casio fx83 / Casio fx85.
Notes	Books, notes or other written material may not be brought into this examination

PLEASE DO NOT REMOVE THIS PAPER FROM THE EXAMINATION ROOM
TURN OVER WHEN INSTRUCTED

—— SOLUTIONS ——

1. Question one is compulsory

a. What is *Pattern Recognition*?

[3 marks]

Answer

Syllabus: Introduction of Pattern Recognition

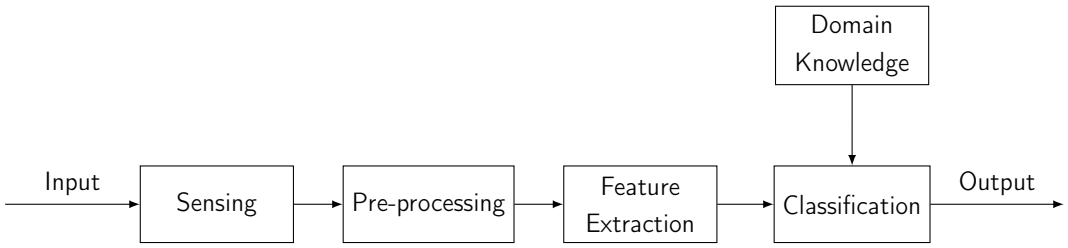
The assignment of a physical object or event to one of several pre-specified categories. 3 marks

b. Draw a block diagram of the key components of a pattern recognition system.

[5 marks]

Answer

Syllabus: Introduction of Pattern Recognition



Marking scheme

1 mark for each block.

c. Sketch a diagram to show a two-dimensional two-class dataset of the following cases. Use *cross* to indicate class 1 and *circle* to indicate class 2.

i. Linearly separable

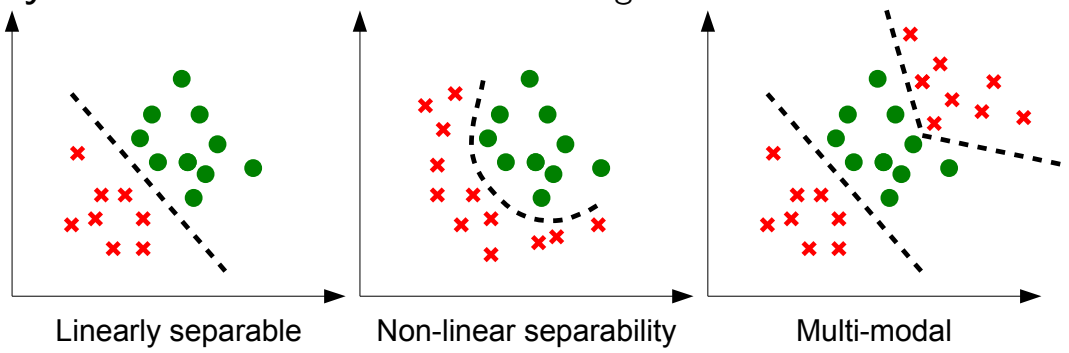
ii. Nonlinearly separable

iii. Multi-modal

[9 marks]

Answer

Syllabus: Introduction of Pattern Recognition



Marking scheme

3 marks for each diagram. In each diagram, the samples are shown by two symbols and the hyperplane should be shown.

d. Given the Minimum Error Rate Classifier for class ω_i with the discriminant function

$$g_i(\mathbf{x}) = \frac{P(\omega_i)}{\sqrt[d/2]{(2\pi)}\sqrt{|\mathbf{\Sigma}_i|}} \exp\big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu_i})^T\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu_i})\big),$$

determine if the following discriminant functions can be used to achieve minimum error rate classification. Explain your answer.

i. $g_i(\mathbf{x}) + 5$

ii. $\frac{25}{3} \times g_i(\mathbf{x})$

iii. $\sin(g_i(\mathbf{x}))$

iv. $-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu_i})^T\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu_i}) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(|\mathbf{\Sigma}_i|) + \ln(P(\omega_i))$

where $\mathbf{x}$ denotes a sample, ω_i denotes class i, $P(\omega_i)$ is the prior probability, $\boldsymbol{\mu_i}$ is the mean vector, $\mathbf{\Sigma}_i$ is the covariance matrix, $\sin$ is the sinusoidal function and $\ln$ is the natural logarithm function.

[8 marks]

Answer

Syllabus: Bayesian Decision Theory and Linear Discriminant Functions

i Yes, adding a constant to discriminant functions does not change the nature of the classifier. 2 marks

ii Yes, multiplying a positive constant to discriminant functions does not change the nature of the classifier. 2 marks

iii No, applying a non-monotonically increasing function to discriminant functions will change the nature of the classifier. $\sin$ is a non-monotonically increasing function. 2 marks

iv Yes, $\ln(g_i(\mathbf{x})) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu_i})^T\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu_i}) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(|\mathbf{\Sigma}_i|) + \ln(P(\omega_i))$. Applying a monotonically increasing function to discriminant functions does not change the nature of the classifier. $\ln$ is a monotonically increasing function. 2 marks

Marking scheme

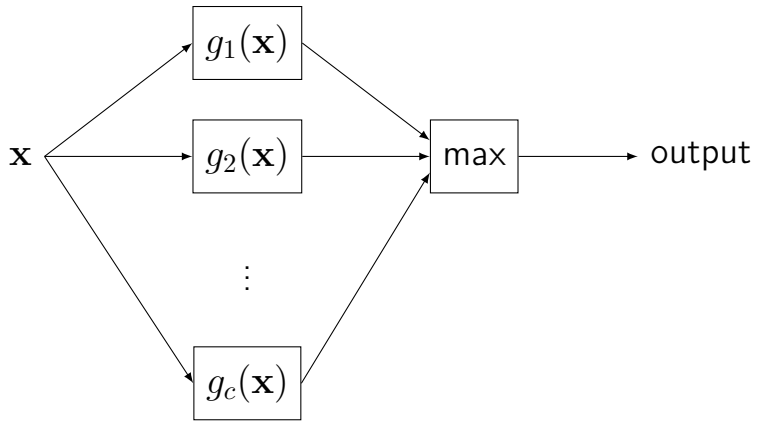
1 mark for yes/no answer and 1 mark for the explanation for each function.

2. a. Draw a block diagram of a *multiclass discriminant function classifier* and explain how it works for classification.

[10 marks]

Answer

Syllabus: Linear Discriminant Functions



6 marks

The multiclass discriminant function classifier evaluates the feature vector $\mathbf{x}$ by the discriminant functions $g_1(\mathbf{x})$, $g_2(\mathbf{x})$, $\cdots$, $g_c(\mathbf{x})$. The $\max$ operator picks the maximum output and assigns the feature vector to class ω_i if $g_j(\mathbf{x}) > g_i(\mathbf{x})\ \forall i \neq j$. 4 marks

Marking scheme

1 mark for each block; 1 mark for "$\mathbf{x}$"; 1 mark for "output".

b. Describe the *Sequential Multiclass Perceptron Learning Algorithm* using pseudocode.

[7 marks]

Answer

Algorithm 1: Sequential Multiclass Perceptron Learning Algorithm

1 **Initialisation:** Define c classes; initialise $\mathbf{a}_1$ to $\mathbf{a}_c$ to arbitrary solution and learning rate η
2 **while** *not converge* **do**
3 **foreach** *exemplar* $(\mathbf{y}, \omega)$ **do**
4 $\omega' = \underset{h}{\operatorname{argmax}}\, g_h(\mathbf{x})$
5 **if** $\omega \neq \omega'$ **then**
6 $\mathbf{a}_\omega \leftarrow \mathbf{a}_\omega + \eta\mathbf{y}$
7 $\mathbf{a}_{\omega'} \leftarrow \mathbf{a}_{\omega'} - \eta\mathbf{y}$
8 **end**
9 **end**
10 **end**

7 marks

Marking scheme

1 mark for each line from lines 1 to 7.

c. Consider the training dataset shown in Table 1.

Class	Feature vector
1	$[-2, 6]$
1	$[-1, -4]$
1	$[3, -1]$
2	$[-3, -2]$
3	$[-4, -5]$

Table 1: Training dataset.

Determine the class of a new feature vector $\mathbf{x} = [-2, 0]$ with the k-nearest-neighbour classifier using Euclidean distance and the following values of k:

i. $k = 1$

ii. $k = 3$

iii. $k = 5$

[8 marks]

Answer

Syllabus: Bayesian Decision Theory and Density Estimation

Class	Feature vector	New feature vector	Euclidean distance
1	$[-2, 6]$	$[-2, 0]$	$\sqrt{(-2-(-2))^2+(6-0)^2} = 6$
1	$[-1, -4]$	$[-2, 0]$	$\sqrt{(-1-(-2))^2+(-4-0)^2} = 4.1231$
1	$[3, -1]$	$[-2, 0]$	$\sqrt{(3-(-2))^2+(-1-0)^2} = 5.0990$
2	$[-3, -2]$	$[-2, 0]$	$\sqrt{(3-(-2))^2+(-2-0)^2} = 2.2361$
3	$[-4, -5]$	$[-2, 0]$	$\sqrt{(-4-(-2))^2+(-5-0)^2} = 5.3852$

5 marks

i When $k = 1$, the new feature vector $[-2, 0]$ is closer to point 4 ($[-3, -2]$). So, it is classified as class 2. 1 mark

ii When $k = 3$, the new feature vector $[-2, 0]$ is closer to points 2, 3 and 5 ($[-1, -4]$, $[3, -1]$ and $[-3, -2]$). As 2 out of 3 points are of class 1, the new feature vector is classified as class 1. 1 mark

iii When $k = 5$, all 5 points are covered. As 3 out of 5 points are of class 1, the new feature vector is classified as class 1. 1 mark

Marking scheme

1 mark for each correct Euclidean distance in the table; 1 mark for each correct classification.

3. a. Name three feature extraction techniques

[3 marks]

Answer

Syllabus: Feature Extraction

- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)
- Independent Component Analysis (ICA)

Marking scheme

1 mark for each correct answer.

b. Consider the dataset in Table 2.

Class	Feature vector $\mathbf{x}^T$
1	$[1, 2]$
1	$[2, 1]$
1	$[3, 3]$
2	$[6, 5]$
2	$[7, 8]$

Table 2: Training dataset.

i. Using *Linear Discriminant Analysis* (LDA), given a projection vector $\mathbf{w}^T = [1, 2]$, find the projection of the feature vectors. Is the dataset in the projection space linearly separable? Explain your answer.

[8 marks]

Answer

Syllabus: Feature Extraction

Class	Feature vector $\mathbf{x}^T$	$y = \mathbf{w}^T\mathbf{x}$
1	$[1, 2]$	$[1, 2] \times [1, 2]^T = 5$
1	$[2, 1]$	$[2, 1] \times [1, 2]^T = 4$
1	$[3, 3]$	$[3, 3] \times [1, 2]^T = 9$
2	$[6, 5]$	$[6, 5] \times [1, 2]^T = 16$
2	$[7, 8]$	$[7, 8] \times [1, 2]^T = 23$

5 marks

It can be seen that the projected dataset is obviously separable into two separate clusters in the projection space, i.e., 5, 4 and 9 for class 1; 16 and 23 for class 2. So the projected dataset in the projection space is linearly separable. 3 marks

Marking scheme

1 mark for each correct answer in the table.

ii. Using *Fisher's method*, determine which of the following projection weights is more effective in the context of *Linear Discriminant Analysis (LDA)*. Explain your answer.

- $\mathbf{w}^T = [-1, 5]$
- $\mathbf{w}^T = [2, -6]$

[14 marks]

Answer

Syllabus: Feature Extraction

Sample mean for class 1 is
$\mathbf{m}_1^T = \frac{1}{3}([1,2] + [2,1] + [3,3]) = [2,2].$ 1 mark

Sample mean for class 2 is
$\mathbf{m}_2^T = \frac{1}{2}([6,5] + [7,8]) = [6.5, 6.5].$ 1 mark

For $\mathbf{w}^T = [-1, 5]$:
Between class scatter (**sb**) $= |\mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)|^2 = |[-1,5] \times ([2,2] - [6.5,6.5])^T|^2 = 324$ 2 marks

Within class scatter (sw) $= \sum_{\mathbf{x} \in \omega_1} (\mathbf{w}^T(\mathbf{x} - \mathbf{m}_1))^2 + \sum_{\mathbf{x} \in \omega_2} (\mathbf{w}^T(\mathbf{x} - \mathbf{m}_2))^2 = ([-1,5] \times ([1,2] - [2,2])^T)^2 + ([-1,5] \times ([2,1] - [2,2])^T)^2 + ([-1,5] \times ([3,3] - [2,2])^T)^2 + ([-1,5] \times ([6,5] - [6.5,6.5])^T)^2 + ([-1,5] \times ([7,8] - [6.5,6.5])^T)^2 = 140$

2 marks

Cost $J(\mathbf{w}) = \frac{sb}{sw} = \frac{324}{140} = 2.3143$ 1 mark

For $\mathbf{w}^T = [2, -6]$:
Between class scatter (**sb**) $= |\mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)|^2 = |[2,-6] \times ([2,2] - [6.5,6.5])^T|^2 = 324$ 2 marks

Within class scatter (sw) $= \sum_{\mathbf{x} \in \omega_1} (\mathbf{w}^T(\mathbf{x} - \mathbf{m}_1))^2 + \sum_{\mathbf{x} \in \omega_2} (\mathbf{w}^T(\mathbf{x} -$

$\mathbf{m}_2))^2 = ([2,-6]\times([1,2]-[2,2])^T)^2+([2,-6]\times([2,1]-[2,2])^T)^2+$
$([2,-6]\times([3,3]-[2,2])^T)^2+([2,-6]\times([6,5]-[6.5,6.5])^T)^2+$
$([2,-6]\times([7,8]-[6.5,6.5])^T)^2 = 184$

2 marks

Cost $J(\mathbf{w}) = \frac{sb}{sw} = \frac{324}{184} = 1.7609$ 1 mark

As $J(\mathbf{w})$ given by $\mathbf{w}^T = [-1,5]$ is higher, it is a more effective projection weight. 2 marks

4. A diagram of 3-layer partially connected neural network is shown in Figure 1.

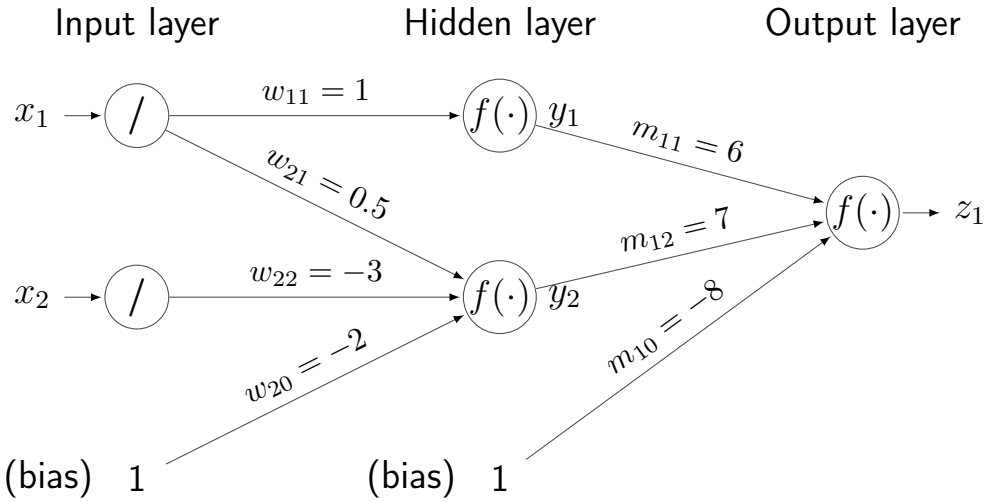


Figure 1: A diagram of 3-layer partially connected neural network.

a. Given that the output of the neural network in Figure 1 can be expressed as

$$z_1 = f\Big(\mathbf{W}_{kj} f\big(\mathbf{W}_{ji}\mathbf{x} + \mathbf{W}_{j0}\big) + \mathbf{W}_{k0}\Big)$$

where $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $f(\cdot)$ is any activation function, determine the matrices $\mathbf{W}_{ji}$, $\mathbf{W}_{kj}$, $\mathbf{W}_{j0}$ and $\mathbf{W}_{k0}$.

[8 marks]

Answer

Syllabus: Multilayer Neural Networks

$\mathbf{W}_{ji} = \begin{bmatrix} 1 & 0 \\ 0.5 & -3 \end{bmatrix}$ 2 marks

$\mathbf{W}_{j0} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$ 2 marks

$\mathbf{W}_{kj} = \begin{bmatrix} 6 & 7 \end{bmatrix}$ 2 marks

$\mathbf{W}_{k0} = \begin{bmatrix} -8 \end{bmatrix}$ 2 marks

b. A classifier is designed as $y = \text{sgn}(z_1)$ where $\text{sgn}(z_1) = \begin{cases} 1 & \text{if } z_1 \geq 0 \\ -1 & \text{if } z_1 < 0 \end{cases}$, $y = 1$ represents class 1 and $y = -1$ represents class 2. Considering a sample $\mathbf{x} = \begin{bmatrix} 2 \\ -6 \end{bmatrix}$ and the activation function $f(\cdot)$ as a logarithmic sigmoid function, i.e., $f(n) = \dfrac{1}{1+e^{-n}}$, determine which class the sample belongs to.

[5 marks]

Answer

Syllabus: Multilayer Neural Networks

Sample: $\mathbf{x} = \begin{bmatrix} 2 \\ -6 \end{bmatrix}$

$$\begin{aligned}
\mathbf{y} &= \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\
&= \text{logsig}\Big(\mathbf{W}_{ji}\mathbf{x} + \mathbf{W}_{j0}\Big) \\
&= \text{logsig}\Big(\begin{bmatrix} 1 & 0 \\ 0.5 & -3 \end{bmatrix}\begin{bmatrix} 2 \\ -6 \end{bmatrix} + \begin{bmatrix} 0 \\ -2 \end{bmatrix}\Big) \\
&= \begin{bmatrix} 0.8808 \\ 1.0000 \end{bmatrix},
\end{aligned}$$

where $\text{logsig}(n) = \frac{1}{1+e^{-n}}$. 2 marks

$$\begin{aligned}
z_1 &= \text{logsig}\Big(\mathbf{W}_{kj}\mathbf{y} + \mathbf{W}_{j0}\Big) \\
&= \text{logsig}\Big(\begin{bmatrix} 6 & 7 \end{bmatrix}\begin{bmatrix} 0.8808 \\ 1.0000 \end{bmatrix} - 8\Big) \\
&= 0.9864
\end{aligned}$$

2 marks

Classifier output: $y = \text{sgn}(z_1) = \text{sgn}(0.9864) = 1$ which indicates that the sample $\mathbf{x} = \begin{bmatrix} 2 \\ -6 \end{bmatrix}$ belongs to class 1. 1 mark

c. Consider the activation function in all the input, hidden and output units as a linear function. The neural network is trained with the stochastic backpropagation algorithm using the cost $J = \frac{1}{2}\|z_1 - t\|^2$ where $\|\cdot\|$ denotes the l_2 (or Euclidean) norm operator. Given the only training sample $\mathbf{x} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$ and its target output $t = 0.8$, determine the updated value of w_{20} at the next iteration using the learning rate $\eta = 0.01$.

[12 marks]

Answer

Syllabus: Multilayer Neural Networks

The update of w_{20} is:

$$\begin{aligned}\Delta w_{20} &= -\eta \frac{\partial J}{\partial w_{20}} \\ &= -\eta \frac{\partial J}{\partial f(net)} \frac{\partial f(net)}{\partial net} \frac{\partial net}{\partial y_2} \frac{\partial y_2}{\partial a} \frac{\partial a}{\partial w_{20}},\end{aligned}$$

2 marks

where

$\eta = 0.01$,

$f(\cdot)$ is a linear function,

$J = \frac{1}{2}\|z_1 - t\|^2$,

$z_1 = f(net) = net = m_{11}y_1 + m_{12}y_2 + m_{10} = 6y_1 + 7y_2 - 8$,

$y_1 = w_{11}x_1 = x_1$,

$y_2 = f(a) = a = w_{21}x_1 + w_{22}x_2 + w_{20} = 0.5x_1 - 3x_2 - 2$.

So, $\frac{\partial J}{\partial f(net)} = \frac{1}{2}\frac{\partial \|z_1 - t\|^2}{\partial z_1} = z_1 - t$, 1 mark

$\frac{\partial f(net)}{\partial net} = 1$, 1 mark

$\frac{\partial net}{\partial y_2} = m_{12}$, 1 mark

$\frac{\partial y_2}{\partial a} = 1$, 1 mark

$\frac{\partial a}{\partial w_{20}} = 1$. 1 mark

Thus, we have $\Delta w_{20} = -\eta(z_1 - t)m_{12}$. 2 marks

Considering $\mathbf{x} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$, we have $y_1 = x_1 = 1$ and $y_2 = 0.5x_1 - 3x_2 - 2 = 0.5 \times 1 - 3 \times 5 - 2 = -16.5$, and $z_1 = 6 \times 1 + 7 \times (-16.5) - 8 = -117.5$. 1 mark

Considering $t = 0.8$, we have $\Delta w_{20} = -\eta(z_1 - t)m_{12} = 0.01 \times (-117.5 - 0.8) \times 7 = 8.2810$. 1 mark

Applying the update rule: $\Delta w_{20} \leftarrow w_{20} + \Delta w_{20} = -2 + 8.2810 = 6.2810$. 1 mark

5. a. What are *Linguistic Hedges* in the context of fuzzy set?

[4 marks]

Answer

Syllabus: Fuzzy Inference System

Linguistic Hedges, like *adverbs* or *objectives*, modify the fuzzy set. The definition of the hedge functions are arbitrary, just to name a few, *very*, *somewhat*, *more or less*, *slightly*. 4 marks

b. Name three categories of *Linguistic Hedges* in the context of fuzzy set? Briefly describe them.

[9 marks]

Answer

Syllabus: Fuzzy Inference System

Three categories of hedges, *concentrations*, *dilations* and *Intensifications*, which change the shape of membership functions. 3 marks

- *concentrations:* It concentrates a fuzzy set by reducing the membership degree of all element. 2 marks
- *dilations:* It stretches or dilates a fuzzy set by increasing the membership degree of all element. (Sometimes referred to as *dilutions*.) 2 marks
- *Intensifications:* It is a combination of *concentration* and *dilation*. It increases (decreases) the membership degree of those elements in the fuzzy set with original membership values greater than 0.5 (less than 0.5). 2 marks

c. Consider a 2-input single-output Sugeno fuzzy inference system with the following 4 rules:

Rule 1: **IF** x is $Small$ and y is $Small$ **THEN** z is $-x+y+1$

Rule 2: **IF** x is $Small$ and y is $Large$ **THEN** z is $-y+3$

Rule 3: **IF** x is $Large$ and y is $Small$ **THEN** z is $-x+3$

Rule 4: **IF** x is $Large$ and y is $Large$ **THEN** z is $x+y+2$

where the membership functions are define as $\mu_{x_{Small}}(x) = 1 - \frac{1}{1+e^{-x}}$, $\mu_{x_{Large}}(x) = \frac{1}{1+e^{-x}}$, $\mu_{y_{Small}}(y) = 1 - \frac{1}{1+e^{-y+5}}$ and $\mu_{y_{Large}}(y) = \frac{1}{1+e^{-y+5}}$.

Given that the fuzzy "AND" operation is the "product" operation, determine the output of the Sugeno fuzzy inference system for $x = 1$ and $y = 8$.

[12 marks]

Answer

Syllabus: Fuzzy Inference System

For $x = 1$ and $y = 8$,

$\mu_{x_{Small}}(1) = 1 - \frac{1}{1+e^{-1}} = 0.2689$, 1 mark

$\mu_{x_{Large}}(1) = \frac{1}{1+e^{-1}} = 0.7311$, 1 mark

$\mu_{y_{Small}}(8) = 1 - \frac{1}{1+e^{-8+5}} = 0.0474$, 1 mark

$\mu_{y_{Large}}(8) = \frac{1}{1+e^{-8+5}} = 0.9526$. 1 mark

Performing fuzzy AND operation using "product" operation, we have

$w_1 = \mu_{x_{Small}}(1) \times \mu_{y_{Small}}(8) = 0.0128$, 1 mark

$w_2 = \mu_{x_{Small}}(1) \times \mu_{y_{Large}}(8) = 0.2562$, 1 mark

$w_3 = \mu_{x_{Large}}(1) \times \mu_{y_{Small}}(8) = 0.0347$, 1 mark

$w_4 = \mu_{x_{Large}}(1) \times \mu_{y_{Large}}(8) = 0.6964$, 1 mark

Individual output of each rule:

$z_1 = -x + y + 1 = -1 + 8 + 1 = 8$,

$z_2 = -y + 3 = -8 + 3 = -5$,

$z_3 = -x + 3 = -1 + 3 = 2$,

$z_4 = x + y + 2 = 1 + 8 + 2 = 11.$

Performing defuzzification, the inferred output is:

$z = \frac{w_1 z_1 + w_2 z_2 + w_3 z_3 + w_4 z_4}{w_1 + w_2 + w_3 + w_2} = \frac{0.0128 \times 8 + 0.2562 \times -5 + 0.0347 \times 2 + 0.6964 \times 11}{0.0128 + 0.2562 + 0.0347 + 0.6964} = 6.5507$

4 marks

6. A support vector machine (SVM) classifier is employed to classify the following samples:

$$\text{Class 1: } \mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

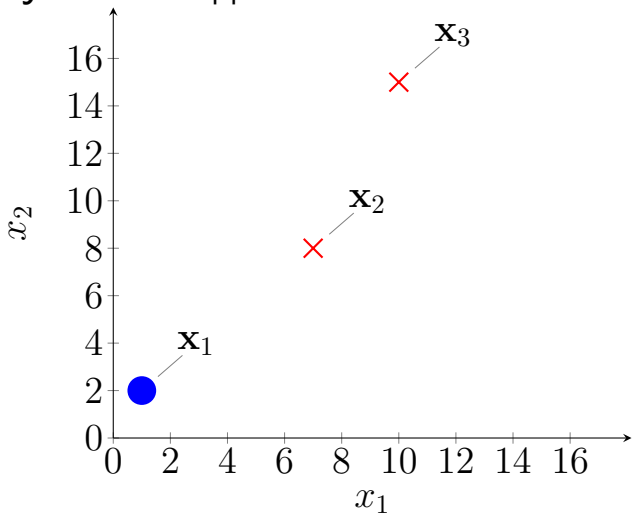
$$\text{Class 2: } \mathbf{x}_2 = \begin{bmatrix} 7 \\ 8 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 10 \\ 15 \end{bmatrix}.$$

a. Identify the support vectors by inspection.

[7 marks]

Answer

Syllabus: Support Vector Machines



3 marks

The samples are linearly separable as a linear hyperplane can be placed in between x_1 and x_2 to separate the samples into two classes. So, $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$ are the support vectors. 4 marks

b. Design an SVM classifier to correctly classify all given samples.

[15 marks]

Answer

Syllabus: Support Vector Machines

Define the label for Class 1 as $+1$ and Class 2 as -1 so we have $y_1 = +1$ for $\mathbf{x}_1$ and $y_2 = y_3 = -1$ for $\mathbf{x}_2$ and $\mathbf{x}_3$. 3 marks

The hyperplane is $\mathbf{w}^T\mathbf{x} + w_0 = 0$. As $\mathbf{x}_1$ and $\mathbf{x}_2$ are support vectors, we have $\mathbf{w} = \lambda_1 y_1 \mathbf{x}_1 + \lambda_2 y_2 \mathbf{x}_2 = \lambda_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \lambda_2 \begin{bmatrix} 7 \\ 8 \end{bmatrix}$. 2 marks

Recall that $y_i(\mathbf{w}^T\mathbf{x} + w_0) = 1$ when $\mathbf{x}$ is a support vector.

For $\mathbf{x} = \mathbf{x}_1$, $\left(\lambda_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \lambda_2 \begin{bmatrix} 7 \\ 8 \end{bmatrix}\right)^T \begin{bmatrix} 1 \\ 2 \end{bmatrix} + w_0 = 1$

$\Rightarrow 5\lambda_1 - 23\lambda_2 + w_0 = 1.$ 2 marks

For $\mathbf{x} = \mathbf{x}_2$, $\left(\lambda_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \lambda_2 \begin{bmatrix} 7 \\ 8 \end{bmatrix}\right)^T \begin{bmatrix} 7 \\ 8 \end{bmatrix} + w_0 = -1$

$\Rightarrow 23\lambda_1 - 113\lambda_2 + w_0 = -1.$ 2 marks

As $\sum_{i=1}^{3} \lambda_i y_i = 0$, we have $\lambda_1 - \lambda_2 = 0 \Rightarrow \lambda_1 = \lambda_2$. So, we have $5\lambda_1 - 23\lambda_2 + w_0 = -18\lambda_1 + w_0 = 1$ and $23\lambda_1 - 113\lambda_2 + w_0 = -90\lambda_1 + w_0 = -1$.

It follows that $\begin{bmatrix} -18 & 1 \\ -90 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ w_0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. 2 marks

It gives $\lambda_1 = \lambda_2 = 0.0278$ and $w_0 = 1.5$. As a result, $\mathbf{w} = 0.0278 \begin{bmatrix} 1 \\ 2 \end{bmatrix} - 0.0278 \begin{bmatrix} 7 \\ 8 \end{bmatrix} = \begin{bmatrix} -0.1668 \\ -0.1668 \end{bmatrix}$ 2 marks

The hyperplane is

$$\mathbf{w}^T\mathbf{x} + w_0 = -0.1668x_1 - 0.1668x_2 + 1.5 = 0.$$

2 marks

c. What is the margin given by the hyperplane obtained in Question 6.b?

[3 marks]

Answer

Syllabus: Support Vector Machines

The hyperplane is $-0.1668x_1 - 0.1668x_2 + 1.5 = 0$, which gives the margin $\frac{2}{\|\mathbf{w}\|} = \frac{2}{\sqrt{(-0.1668)^2+(-0.1668)^2}} = 8.4785.$ 3 marks.