Index no. | | | | | | | |

UNIVERSITY OF MORATUWA, SRI LANKA
Faculty of Engineering
Department of Electronic & Telecommunication Engineering
B.Sc. Engineering
Semester 5 Quiz

EN 3150—Pattern Recognition

Time Allowed: 1 hour   Sep. 2024

INSTRUCTIONS TO CANDIDATES

- This quiz contains 15 MCQ questions on 11 pages.
- Answer **all** the questions.
- This is a open-book examination.

Q1. A data set of four data samples are given in table. Suppose that regression model $f(\boldsymbol{x}) = w_0 + w_1(x_1 x_2)$ fits to the given data set. Use all the data samples to find parameters of the regression model (w_0 and w_1). [**10 marks**]

Sample index (i)	$x_{1,i}$	$x_{2,i}$	y_i
1	$\frac{1}{5}$	5	5.5
2	3	$\frac{2}{3}$	8
3	9	$\frac{1}{3}$	11.5
4	8	$\frac{1}{2}$	15

$w_0 : 2.0$ and $w_1 : 3.2$

Q2. Consider a dataset consisting of three flower classes: Iris Setosa, Iris Versicolor, and Iris Virginica. In this dataset each data sample is represented by a 4-dimensional vector. After the learning process, a linear classifier is given as follows:

$$\boldsymbol{W} = \begin{bmatrix} 0.41 & 1.46 & -2.26 & -1.02 \\ 0.42 & -1.61 & 0.57 & -1.40 \\ -1.70 & -1.53 & 2.47 & 2.55 \end{bmatrix}^T, \text{ and } \boldsymbol{b} = \begin{bmatrix} 0.26 & 1.09 & -1.21 \end{bmatrix}^T.$$

Suppose that two data samples $\boldsymbol{x}_1$ and $\boldsymbol{x}_2 \in \mathbb{R}^{4\times 1}$ are fed to this linear classifier. $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are given by

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 4 & 1 \\ 8 & 1 & 3 & 2 \end{bmatrix}^T \in \mathbb{R}^{4\times 2}.$$

Output of the linear classifier for $i-$th data sample is given by $\boldsymbol{y}_i = f\left(\boldsymbol{W}^T \boldsymbol{x}_i + \boldsymbol{b}\right) \in \mathbb{R}^{3\times 1}$. Here, $f(\cdot)$ is the sigmoid function. Suppose that probability of $i-$th data sample belongs to $j-$th class denoted by $P_{j,i}$. What are the probabilities of both data samples? [**10 marks**]

Index (i)	$P_{1,i}$	$P_{2,i}$	$P_{3,i}$
1	0.0312	0.3288	0.64
2	0.024	0.9523	0.0238
Index (i)	$P_{1,i}$	$P_{2,i}$	$P_{3,i}$
1	0.0223	0.2351	0.4576
2	0.0215	0.852	0.0212
Index (i)	$P_{1,i}$	$P_{2,i}$	$P_{3,i}$
1	0.4	0.2	0.4
2	0.3	0.2	0.5
Index (i)	$P_{1,i}$	$P_{2,i}$	$P_{3,i}$
1	0.0879	0.0698	0.6814
2	0.001	0.8146	0.8849

Q3. Consider the linear regression model of $y(\boldsymbol{x}_i) = w_0 + w_1 x_{1,i} + w_2 x_{2,i}$ and loss function with ℓ_2 regularization as

$$L(\boldsymbol{w}) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\left(y_i - y(\boldsymbol{x_i})\right)^2 + \lambda\|\boldsymbol{w}\|_2^2.$$

Suppose that for the j-th iteration $w_0^{(j)}$, $w_1^{(j)}$ and $w_2^{(j)}$ are given by 1, 0.5, 1.5, respectively. Find the value of w_2 for (j+1)-th iteration using stochastic gradient descent (SDG) algorithm with a learning rate of 0.1 and regularization parameter $\lambda = 0.1$. Here, the data sample is used for the SGD is $y_i = 6$, $x_{1,i} = 2$, and $x_{2,i} = 2$, respectively. [**10 marks**]

(a) 1.67

(b) 1.7

(c) 1.55

(d) 1.9

Index (i)	1	2	3	4	5	6	7	8	9	10
$x_{1,i}$	4	2	-5	3	3	4	-11	-8	-13	-7
$x_{2,i}$	1	-3	7	4	8	6	-6	-14	-10	-13
Class label y_i	c_1	c_1	c_1	c_1	c_1	c_1	c_2	c_2	c_2	c_2

Q4. Table displays feature values x_1 and x_2 of data samples which are belongs to two classes namely c_1 and c_2. Note that feature values are rounded to nearest integer. Here, it is assumed that class-conditional densities are Gaussian distributed, i.e., $p(\boldsymbol{x}|y = c_k, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$. Calculate mean vectors $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and prior probabilities of classes ($p(y = c_1|\boldsymbol{\theta}) = \pi_1$ and $p(y = c_2|\boldsymbol{\theta}) = \pi_2$). [**10 marks**]

(a) $\boldsymbol{\mu}_1 = [1.83\ 3.83]^T$, $\boldsymbol{\mu}_2 = [-9.75\ -10.75]^T$, $\pi_1 = 0.6$ and $\pi_2 = 0.4$

(b) $\boldsymbol{\mu}_1 = [1.83\ 3.83]^T$, $\boldsymbol{\mu}_2 = [-9.75\ -10.75]^T$, $\pi_1 = 0.5$ and $\pi_2 = 0.5$

(c) $\boldsymbol{\mu}_1 = [1.83\ 3.83]^T$, $\boldsymbol{\mu}_2 = [-9.75\ -10.75]^T$, $\pi_1 = 0.4$ and $\pi_2 = 0.6$

(d) $\boldsymbol{\mu}_1 = [[4.33\ 3.0]^T$, $\boldsymbol{\mu}_2 = [-12.25\ -11.0]^T$, $\pi_1 = 0.6$ and $\pi_2 = 0.4$

Q5. Consider the linear model given by $z_i = w_0 + w_1 x_i$. where $w_1 = 2$ and $w_0 = -5$. The output of this linear model is then mapped to a probabilistic output using the function below:

$$p\left(y_i = \text{Class A}\middle|x_i, \boldsymbol{w}\right) = \left(\frac{f(z_i) + 1}{2}\right), \text{ with } f(z_i) = \left(\frac{e^{z_i} - e^{-z_i}}{e^{z_i} + e^{-z_i}}\right).$$

Here, y_i is the class label of $i-$th data sample and there are only two possible classes to which a data sample can belong: Class A and Class B.
Choose correct answer for the input value $x_i = 1.5$. [**10 marks**]

(a) The probability that x_i belongs to class B is 0.9820

(b) The probability that x_i belongs to class B is 0.0180

(c) The probability that x_i belongs to class B is 0.1192

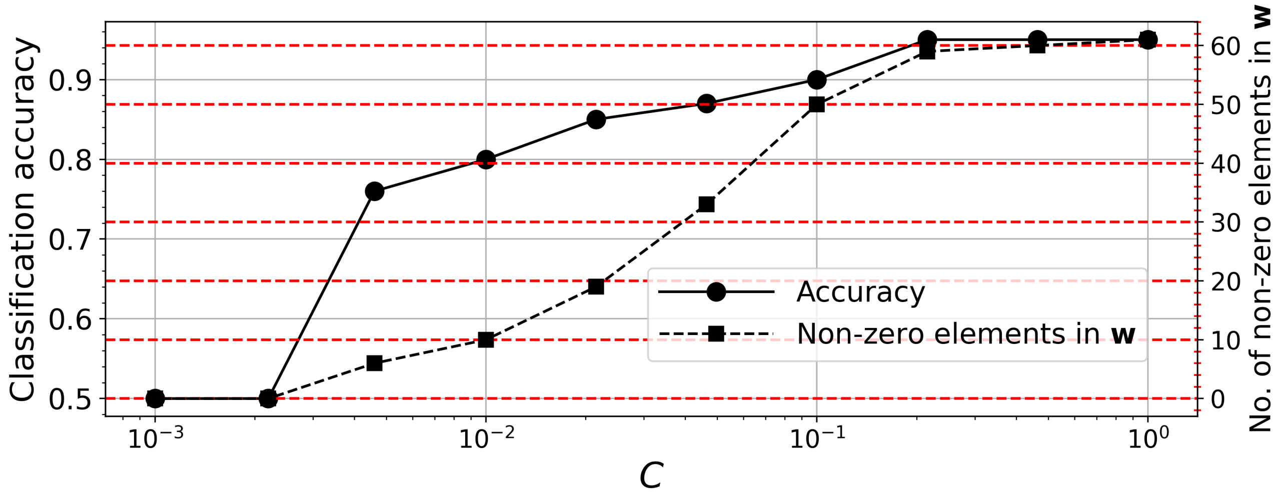
(d) The probability that x_i belongs to class B is 0.8808

(e) The probability that x_i belongs to class A is 0.1192

Q6. Binary class logistic regression with a regularization term, minimizes the following cost function

$$\mathscr{L}(\boldsymbol{w}) = C\left(\sum_{i=1}^{N}\left[-y_i \log(\mu_i) - (1-y_i)\log(1-\mu_i)\right]\right) + \|\boldsymbol{w}\|_1.$$

Here, $\mu_i = \text{sigm}(\boldsymbol{w}^T \boldsymbol{x}_i)$ with features $\boldsymbol{x}_i \in \mathbb{R}^{64\times 1}$, where sigm$(\cdot)$ is the sigmoid function. Figure illustrates the classification accuracy and the number of non-zero elements in the coefficient vector ($\boldsymbol{w}$) with respect to different C values.
What is the percentage increase in features ($\boldsymbol{x}$) required to change classification accuracy from 80% to 90%? **[5 marks]**



(a) 62.50%

(b) 66.66%

(c) 33.33%

(d) 75.00%

Model	Bias	Variance	Irreducible error
Model 1	-0.5	0.5	0.1
Model 2	0.6	0.3	0.1

Q7. Table provides the bias, variance, and irreducible error of two machine learning models for a given data set. Based on this information, which model would you choose? [**05 marks**]

Model 1
Model 2
Both Model 1 and 2
Cannot be determined

Q8. Suppose you have a huge data set and you have time constrains. Here, which variation of the gradient descent algorithm would you select ?

(a) Stochastic Gradient Descent

(b) Batch Gradient Descent

[**05 marks**]

	Email spam detection				Credit card fraud detection			
	True Pos.(TP)	False Pos.(FP)	True Neg.(TN)	False Neg.(FN)	TP	FP	TN	FN
Method A	960	40	940	60	900	100	850	150
Method B	970	30	930	70	850	150	900	100

Q9. You have two machine learning algorithms, namely method A and method B, which are trained and tested on two scenarios: credit card fraud detection and email spam detection. Calculate recall and precision values for each scenario based on the data given in the Table. Based on these results, which machine learning method would you choose for each scenario? [**05 marks**]

(a) Credit card fraud detection: Method B with precision=0.85 and recall=0.89, Email spam detection: Method B with precision=0.97 and recall=0.93

(b) Credit card fraud detection: Method B with precision=0.75 and recall=0.79, Email spam detection: Method B with precision=0.77 and recall=0.63

(c) Credit card fraud detection: Method A with precision=0.9 and recall=0.85, Email spam detection: Method A with precision=0.96 and recall=0.94

(d) Credit card fraud detection: Method B with precision=0.85 and recall=0.89, Email spam detection: Method A with precision=0.96 and recall=0.94

(e) Credit card fraud detection: Method A with precision=0.9 and recall=0.85, Email spam detection: Method B with precision=0.97 and recall=0.93

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$$

Answer:

- Credit card fraud detection: Here, higher recall is more important (missing fraudulent (False Negatives) activities can result in significant financial losses).
- Email spam detection: Here, higher precision is more important. False positives can lead to important messages being moved to spam and user may missed them.

Q10. Advantages of generative classifiers over discriminative classifiers? [**05 marks**]

(a) Generative classifiers can be used with with missing data/unlabeled data

(b) Generative classifiers can be used to generate new samples from the learned distribution of the data

(c) Generative classifiers always achieve better accuracy

(d) Generative classifiers always faster to train

Q11. Suppose that both dependent and independent variables have measurement errors. Which algorithm would you choose? [**05 marks**]

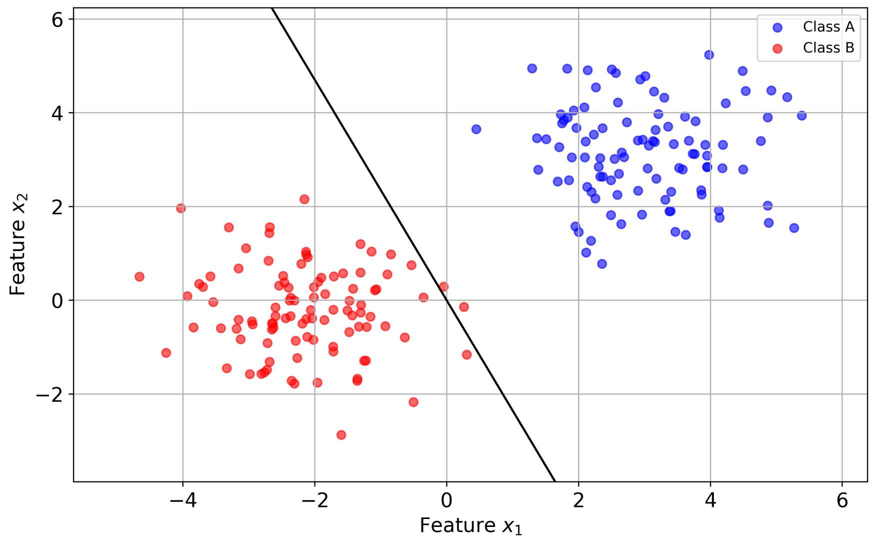
(a) Total least squares (TLS)
(b) Ordinary least-squares (OLS)
(c) Ridge Regression
(d) Polynomial Regression
(e) Least Absolute Shrinkage and Selection Operator (LASSO)
(f) None of the above

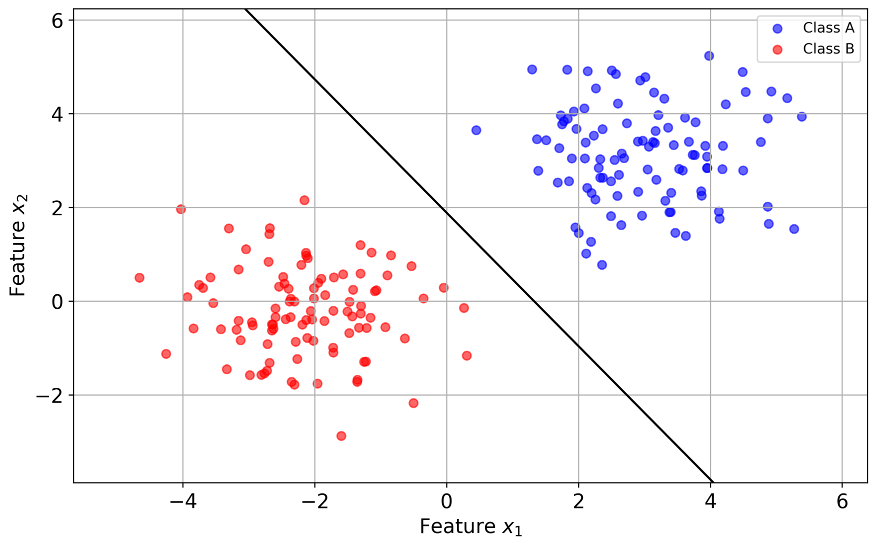
Q12. Binary class logistic regression with a regularization term, minimizes the following custom cost function

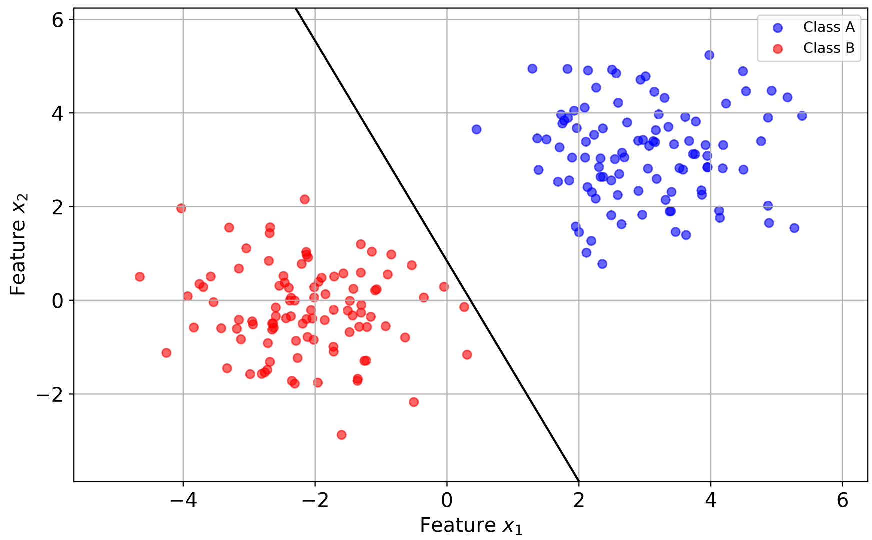
$$\mathscr{L}(\boldsymbol{w}) = \sum_{i=1}^{N} \left[-y_i \log(\mu_i) - (1-y_i)\log(1-\mu_i)\right] + \lambda w_0^2.$$

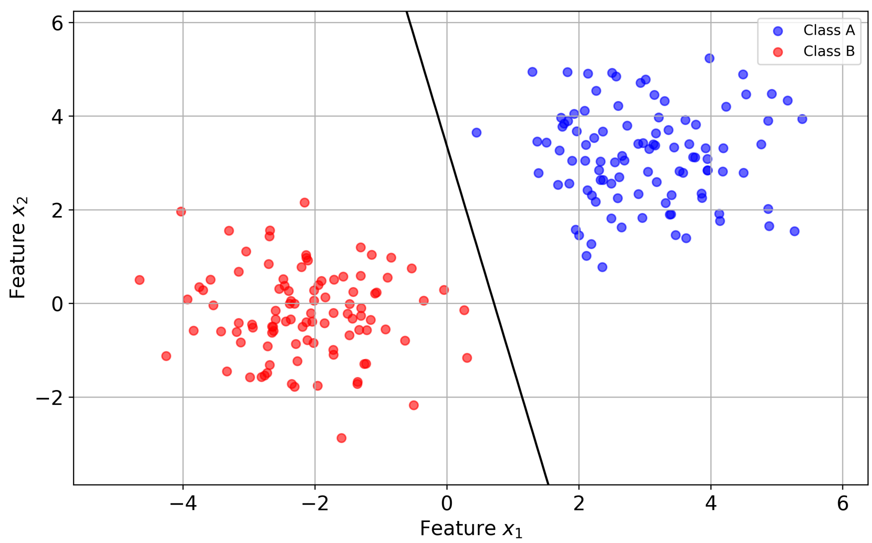
Here, $\mu_i = \text{sigm}(w_0 + w_1 x_{1,i} + w_2 x_{2,i})$, where sigm$(\cdot)$ is the sigmoid function and λ is the regularization parameter. Suppose λ is a very large number, i.e., $\lambda \to \infty$. Choose a possible decision boundary represented by the black line in the figure, which results from minimizing the loss function $\mathscr{L}(\boldsymbol{w})$.

[**05 marks**]









Q13. Consider the following data set ($\boldsymbol{X}$), which consists of five samples of ten distinct features. Which scaling method is appropriate for preserving the structure of this data set? [**05 marks**]

$$\boldsymbol{X} = \begin{bmatrix} 0 & 0 & 0.5 & 0 & 200 & 3 & 0 & 0.1 & 0 & -100 \\ 4 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.75 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \\ 0.1 & 0 & 0 & 0 & 2 & 12 & 0 & -1000 & 1000 & 0 \end{bmatrix}$$

(a) Max-abs scaling

(b) Min-max scaling

(c) Min-max scaling

(d) Standard scaling

(e) Robust scaling

Q14. Figure shows data samples of three distinct classes. Here, it is assumed that class-conditional densities are Gaussian distributed. Which of the following statements are true? [**05 marks**]

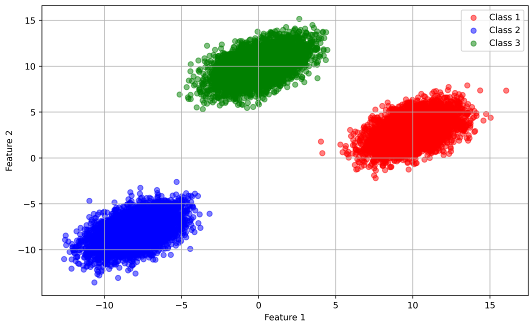


Figure A

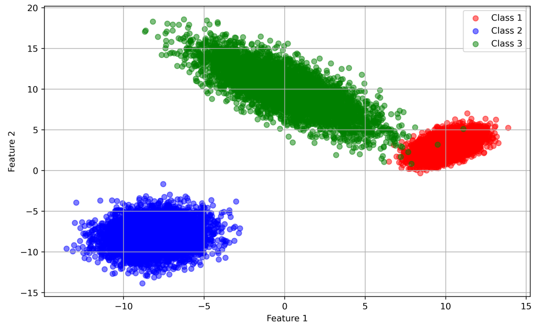


Figure B

(a) Figure A: all three classes have same covariance matrix and different mean vectors.

(b) Figure B: all three classes have different covariance matrices and different mean vectors.

(c) Figure A: all three classes have different covariance matrices and different mean vectors.

(d) Figure B: all three classes have different covariance matrices and same mean vectors.

Q15. A simple model will have low bias and low variance, and a complex model will have high bias and high variance. [**05 marks**]

(a) False

(b) True