Sample Question Format
(For all courses having end semester Full Mark=50)

KIIT Deemed to be University
Online End Semester Examination(Spring Semester-2021)

Subject Name & Code: Machine Learning CS3035
Applicable to Courses: B.Tech (IT)

Full Marks=50 **Time:2 Hours**

SECTION-A (Answer All Questions. Each question carries 2 Marks)

Time:30 Minutes **(7×2=14 Marks)**

Question No	Question Type (MCQ/SAT)	Question	CO Mapping	Answer Key (For MCQ Questions only)
Q.No:1	MCQ	Which of the following outcome is odd man out in the below list? A) R Squared B) RMSE C) Kappa D) All of the mentioned	1	C
		Point out the wrong statement. A) Regression through the origin yields an equivalent slope if you center the data first B) Normalizing variables results in the slope being the correlation C) Least squares is not an estimation tool D) None of the mentioned	1	C
		Which of the following implies no relationship with respect to correlation? A) Cor(X, Y) = 0 B) Cor(X, Y) = 1 C) Cor(X, Y) = 2 D) All of the mentioned	1	A
		Residual _____ plots investigate normality of the errors. A) RR B) QQ C) PP D) None of the mentioned	1	B
Q.No:2	MCQ	Which of the following is the correct formula for total variation? A) Total Variation = Residual Variation – Regression Variation	2	B

		B) Total Variation = Residual Variation + Regression Variation C) Total Variation = Residual Variation * Regression Variation D) All of the mentioned		
		Which of the following is required by K-means clustering? A) defined distance metric B) number of clusters C) initial guess as to cluster centroids D) All of these	2	D
		Minimizing the likelihood is the same as maximizing -2 log likelihood. A) True B) False	2	A
		Let us say that we have computed the gradient of our cost function and stored it in a vector g. What is the cost of one gradient descent update given the gradient? (A) O(N) (B) O(D) (C) O(ND) (D) O(ND2)	2	B
Q.No:3	MCQ	Consider a linear-regression model with N = 3 and D = 1 with input-ouput pairs as follows: y1 = 22, x1 = 1, y2 = 3, x2 = 1, y3 = 3, x3 = 2. What is the gradient of mean-square error (MSE) with respect to β1 when β0 = 0 and β1 = 1? Give your answer correct to two decimal digits. A) -1.66 B) 1.66 C) 1.39 D) None of these	3	A
		Gradient of a continuous and differentiable function (A) is zero at a minimum (B) is non-zero at a maximum (C) decreases as you get closer to the minimum (D) All of these	3	D
		Which of the following sentence is FALSE regarding regression? (A) It relates inputs to outputs. (B) It is used for prediction. (C) It may be used for interpretation. (D) It discovers causal	3	D

		relationships		
		For the one-parameter model, mean-Square error (MSE) is defined as follows: $\frac{1}{2N}\sum_{i=1}^{N}(y_n - \beta_0)^2)$. We have a half term in the front because, (A) scaling MSE by half makes gradient descent converge faster. (B) presence of half makes it easy to do grid search. (C) it does not matter whether half is there or not. (D) none of the above	3	C
Q.No:4	MCQ	Lasso can be interpreted as least-squares linear regression where (A) weights are regularized with the L1 norm (B) weights are regularized with the L2 norm (C) the weights have a Gaussian prior (D) the solution algorithm is simpler	3	A
		Regarding bias and variance, which of the follwing statements are true? (Here 'high' and 'low' are relative to the ideal model.) (A) Models which overfit have a high bias. (B) Models which overfit have a low bias. (C) Models which underfit have a high variance. (D) All of these	3	B
		In K-fold cross-validation, K is (A) A float (decimal) value (B) An integer value (C) An complex imaginary value (D) None of these	3	B
		The second principal (PC2) component is _____ to the first principal component (PC1) A) Orthogonal B) Inverse C) Transpose D) None of these	3	A
Q.No:5	MCQ	What is the cosine similarity between [4, 3, 3, 5], and [2, 0,	1	A

		0, 0]?. A) 0.52 B) 0.62 C) 0.72 D) 0.74		
		Find the odd out from the following list: Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Stochastic Gradient Descent (SGD), and Gravitational Search Algorithm (GSA). A) GA B) PSO C) SGD D) None of these	1	C
		What is the cosine similarity between [5, 2, 0, 5], and [2, 0, 0, 0] A) 0.58 B) 0.55 C) 0.75 D) 0.68	1	D
		The Manhattan distance between two points (10, 10) and (30,30) is: A) 20 B) 30 C) 40 D) 50	1	C
Q.No:6	MCQ	Suppose we train a hard-margin linear SVM on n > 100 data points in , yielding a hyperplane with exactly 2 support vectors. If we add one more data point and retrain the classifier, what is the maximum possible number of support vectors for the new hyperplane (assuming the n + 1 points are linearly separable)? (A)2 (B)3 (C)(n+1) (D)n	5	C
		A valid Kernel function follows _____ condition. A) Vornoi B) Minkowiski C) Mercers D) None of them	5	C
		In Soft-margin Support Vector Classifier, the _____ variable allows few instances within the margin. A) Slack B) Sigma C) Beta	5	B

		D) None of these		
		_____ multipliers used to integrate the inequality constraints into the main objective function. A) Euler B) Lagrangian C) Lypanov D) None of these	5	B
Q.No:7	MCQ	Neural networks (A) optimize a convex cost function (B) can be used for regression as well as classification (C) always output values between 0 and 1 (D) None of these	6	B
		Sigmoidal activation function maps the neuron output between A) -1 to +1 B) 0 to 1 C) Either 0 or 1 D) None of these	6	B
		A single input and output nuron has an input of 2, a weight of 6 and a bias of -5.5. What will be the output if the activation function is bipolar sigmoid (where, slope s=0.6 and round off 4 decimal places) A) 0.9802 B) 0.7806 C) 0.9881 D) None of these	6	A
		In a multi-layer neural network, the term “multi” suggests A) Only one hidden layer in between input and output layers B) One or more hidden layer(s) in between input and output layers C) Always more than one hidden layers in between input and output layers D) None of these	6	B

SECTION-B(Answer Any Three Questions. Each Question carries 12 Marks)

Time: 1 Hour and 30 Minutes **(3×12=36 Marks)**

<table>
<tr><th>Question No</th><th>Question</th><th>CO Mapping (Each question should be from the same CO(s))</th></tr>
<tr><td>Q.No:8</td><td>
A) Why do we call Naive Bayes classifier “Naive”? State its advantages and disadvantages. [1+2=3]
B) Explain the following with examples: One-Against-All (OAA) and One-Against-One (OAO). [2+2=4]
C) Derive the Naive Bayes classification algorithm using the following toy dataset. [5]
<table>
<tr><th>Example No.</th><th>Color</th><th>Type</th><th>Origin</th><th>Stolen?</th></tr>
<tr><td>1</td><td>Red</td><td>Sports</td><td>Domestic</td><td>Yes</td></tr>
<tr><td>2</td><td>Red</td><td>Sports</td><td>Domestic</td><td>No</td></tr>
<tr><td>3</td><td>Red</td><td>Sports</td><td>Domestic</td><td>Yes</td></tr>
<tr><td>4</td><td>Yellow</td><td>Sports</td><td>Domestic</td><td>No</td></tr>
<tr><td>5</td><td>Yellow</td><td>Sports</td><td>Imported</td><td>Yes</td></tr>
<tr><td>6</td><td>Yellow</td><td>SUV</td><td>Imported</td><td>No</td></tr>
<tr><td>7</td><td>Yellow</td><td>SUV</td><td>Imported</td><td>Yes</td></tr>
<tr><td>8</td><td>Yellow</td><td>SUV</td><td>Domestic</td><td>No</td></tr>
<tr><td>9</td><td>Red</td><td>SUV</td><td>Imported</td><td>No</td></tr>
<tr><td>10</td><td>Red</td><td>Sports</td><td>Imported</td><td>Yes</td></tr>
</table>
Classify a “Red SUV Domestic” using Naive Bayes Classifier.
</td><td>CO3</td></tr>
<tr><td></td><td>
A) What is slack variable and its role in soft-margin SVM classifier? Explain the concept using suitable diagram. [2]
B) Find the entropy, information gain and draw the decision trees for the following set of training dataset. [4]
<table>
<tr><th>Gender</th><th>Car ownership</th><th>Travel cost</th><th>Income Level</th><th>Transportation (Class)</th></tr>
<tr><td>Male</td><td>0</td><td>Cheap</td><td>Low</td><td>Bus</td></tr>
<tr><td>Male</td><td>1</td><td>Cheap</td><td>Medium</td><td>Bus</td></tr>
<tr><td>Female</td><td>1</td><td>Cheap</td><td>Medium</td><td>Train</td></tr>
<tr><td>Female</td><td>0</td><td>Cheap</td><td>Low</td><td>Bus</td></tr>
<tr><td>Male</td><td>1</td><td>Cheap</td><td>Medium</td><td>Bus</td></tr>
<tr><td>Male</td><td>0</td><td>Standard</td><td>Medium</td><td>Train</td></tr>
<tr><td>Female</td><td>1</td><td>Standard</td><td>Medium</td><td>Train</td></tr>
<tr><td>Female</td><td>1</td><td>Expensive</td><td>High</td><td>Car</td></tr>
<tr><td>Male</td><td>2</td><td>Expensive</td><td>Medium</td><td>Car</td></tr>
<tr><td>Female</td><td>2</td><td>Expensive</td><td>High</td><td>Car</td></tr>
</table>
C) Explain the following terms with examples: Confusion Matrix, Accuracy, Precision, Recall, F1-score and Area Under the Curve (AUC). [6]
</td><td></td></tr>
<tr><td></td><td>
A) Explain the following with examples: One-Against-All (OAA) and One-Against-One (OAO). [4]
B) States the Mercers conditions for a valid kernel function. Name at least four valid kernel functions used in SVM
</td><td></td></tr>
</table>

<table>
<tr><td></td><td>classifier with appropriate equations. [4]
C) The merits and demerits of the SVM classifier. [4]</td><td></td></tr>
<tr><td><u>Q.No:9</u></td><td>A) What is “curse of dimensionality”? State the possible remedies. [2+1=3]
B) What is a Covariance Matrix? State its limitation. [2+1=3]
C) Using PCA, reduce the Dimension of the given data from 2 to 1. [6]

<table><tr><td>Feature /Example</td><td>Ex1</td><td>Ex2</td><td>Ex3</td><td>Ex4</td></tr><tr><td>F1</td><td>12</td><td>4</td><td>8</td><td>17</td></tr><tr><td>F2</td><td>5</td><td>9</td><td>16</td><td>7</td></tr></table>
A) Differentiate between lossy and lossless feature/attribute reduction with suitable examples. [3]
B) What is a Covariance Matrix? State its limitation. [2+1=3]
C) Explain the Principal Component Analysis (PCA) and reduce the following dataset step-by-step from 2 dimensions to 1. [6]
<table><tr><td>Feature /Example</td><td>Ex1</td><td>Ex2</td><td>Ex3</td><td>Ex4</td></tr><tr><td>F1</td><td>6</td><td>-3</td><td>-2</td><td>7</td></tr><tr><td>F2</td><td>-4</td><td>5</td><td>6</td><td>-3</td></tr></table>
A) Differentiate between lossy and lossless feature/attribute reduction with suitable examples. [2]
B) What is a principal component? Explain the (mathematical) relationship between the first and the second principal components. [2+2=4]
C) Using PCA, reduce the Dimension of the given data from 2 to 1. [6]

<table><tr><td>Feature /Example</td><td>Ex1</td><td>Ex2</td><td>Ex3</td><td>Ex4</td></tr><tr><td>F1</td><td>12</td><td>4</td><td>8</td><td>17</td></tr><tr><td>F2</td><td>5</td><td>9</td><td>16</td><td>7</td></tr></table></td><td>CO5</td></tr>
<tr><td><u>Q.No:10</u></td><td>A) What are over-fitting and under-fitting? Explain them with suitable examples. [4]
B) Explain the Elbow technique to determine sn appropriate “K” value in K-NN classifier. [2]
C) What are the intra-cluster and inter-cluster distances? Using K-means Clustering Algorithm, cluster the following data points: P1=(75,102), P2=(201,16), P3=(68, 80), P4=(188,36), P5=(165,55) and P6=(100,42). where, K=2, and use Euclidean distance for the purpose. (start the computation with P2 and P3 as the two initial centroids points) [1+5=6]

A) What is the significance of bias in the decision boundary? Explain the Bias-Variance relationship in machine learning with suitable diagrams. [2+2=4]
B) Why does we use the term “regression” in the Logistic Regression classification algorithm. [2]
C) Using KNN algorithm and the given data set, predict the class label of the test data point (16,8), where K=3 and Euclidean distance. [6]
X Y Label
----
10 05 0</td><td>CO2, CO5</td></tr>
</table>

	6.5 11 1 7 15 1 12 05 0 8 10 1 15 8 0	
	A. What are the over-fitting and the under-fitting? Explain them with suitable examples. [3] B. What are the intra-cluster and inter-cluster distances? Explain the Elbow technique to determine an appropriate "K" value in K-NN classifier. [1.5+1.5=3] C. Using K-means Clustering Algorithm, cluster the following data points: P1=(75,102), P2=(201,16), P3=(68, 80), P4=(188,36), P5=(165,55) and P6=(100,42). where, K=2, and use Euclidean distance for the purpose. (*start the computation with P1 and P3 as the two initial centroids points*) [6]	
Q.No:11	A) What is Binary Activated Neuron model? States its limitations. [4] B) A two input single output neuron model has weights value [1.5 -2.1] and bias of -3. It is given an input [2.0 2.5]. What will be the output if the binary step function threshold=1 is used? [4] C) Describe Multi-layer Perceptron Neural Network with suitable mathematical expressions and diagram. [4]	CO6
	A) What is perceptron? Explain it with an example. [4] B) Differentiate between linearly and non-linearly separable datasets. [2] C) Solve XOR problem with a two input artificial neuron model. [6]	
	A) What are the learning rate and the momentum in Artificial Neural Network (ANN) model? State different learning rules used in ANN. [2+2=4] B) Draw a diagram of a multiple input single output artificial neural network and compute its input-output relationship. [4] C) Describe the Backpropagation algorithm using appropriate mathematical expressions. [4]	