



BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
WORK INTEGRATED LEARNING PROGRAMMES DIVISION
POST GRADUATE PROGRAMME IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING
COMPREHENSIVE EXAM – CLASSIFICATION REGULAR

Course Code : PCAM ZC311 Course Title : Classification
Marks : 40
Date of Exam :

Q1 [2+4+2=8 Marks]

- a) Explain how probabilistic generative models are different from probabilistic discriminative models using example(s). [2 marks]
- b) Given below is the confusion matrix for multi-class classifier when it was run on test data.

	Actual Class →	Class A	Class B	Class C
Predicted Class	Class A	23	34	22
	Class B	17	16	8
	Class C	20	10	10

Comment on the performance of this classifier by calculating precision and recall with respect to Class B. [4 marks]

- c) Do we need feature scaling in k-NN algorithm? Justify your answer using an example. [2 marks]

Q 2 [5+5=10 Marks]

- a) Suppose you are working in ABC organization. As part of profile creation, the firm captures various important details about its members. Some of the samples are shown in the below table. Using this sample dataset, with the help of Naïve Bayesian classification technique, classify the following tuple either as “High”, or “Middle” or “Low” income bracket member. **Apply “Laplacian” correction to avoid zero probability.** [5 marks]
- {SrNo = “7”, Age = “Above_40”, Gender = “F”, Occupation = “Data Scientist”}

Sr No	Age	Gender	Occupation	Income Bracket
1	Above_40	F	Software Engineer	High
2	Below_30	M	Marketing Executive	Middle
3	Between_31_to_40	M	Unemployed	Low
4	Below_30	M	Data Scientist	High
5	Between_31_to_40	F	Software Engineer	High
6	Below_30	F	Unemployed	Low

- b) Can the cost function which is used in linear regression work in logistic regression? Justify your answer [2.5 marks]
- c) How many binary classifiers are required to use logistic regression for 4 classes using one-vs-all approach? Explain how does it work? [2.5 marks]

Q 3

- a) Consider the following training data set (with three attributes viz., Income, Education, Gender and class/target variable is “Buys Gadget”) for a binary class problem. The attributes are nominal with two possible values. We intend to create a decision tree model using Information Gain. Which attribute would the decision tree induction algorithm choose for the root node? (3 marks)

Income	Education	Gender	Buys Gadget
High	Low	Male	Yes
Low	High	Female	No
High	Low	Male	Yes
High	High	Male	Yes
Low	Low	Male	No
High	Low	Female	No
Low	High	Male	No
Low	Low	Male	No
High	Low	Female	No
High	High	Male	Yes

- b) Can Information gain always help you get the most optimal decision tree? Justify your answer [2 mark]

Q 4. [5+4+4=12 Marks]

- a) How is SVM different from logistic regression? Why SVM algorithm is called a maximum margin classifier? [2+3=5 marks]
- b) SVM is a sparse kernel machine, justify using KKT conditions. [4 marks]
- c) Give any example of non-linearly separable data. Can SVM be used for nonlinearly separable data? If Yes, what is the approach used? If No, how is nonlinear data handled? [4 marks]

Q 5. [2+3=5 Marks]

- a) What are the techniques used by Random forest algorithm for feature selection for building different decision trees? [2 mark]
- b) Which of the two ensemble learners, bagging or boosting, would you expect to be more robust to noise in the data? [1 mark]
- c) The error rate of the ensemble classifier on the original training data is 0. In boosting would you stop your iteration? Why or why not? [2 mark]
