KTUPredict practice paper, Set 1: built from the most repeated past questions. Not an official KTU paper.

PCCST503 Machine Learning

Practice paper · maximum marks 60

PART A Answer all questions. Each carries 3 marks. (24 marks)

1.	Differentiate between regression and classification.	**3**
2.	A coin gives 7 heads in 10 tosses, and the prior is Beta(2, 2). Find the MAP estimate.	**3**
3.	b) A bank labels loan applicants as default (1) or not (0). Four applicants have debt-ratio x and label: (1, 1), (2, 0), (3, 1), (4, 1). Starting from $\theta 0 = \theta 1 = 0$ with $\alpha = 1$, do one batch gradient descent step for logistic regression and compare the costs. (3) (Figure: θ subscripts inferred)	**3**
4.	c) Show that MAP estimation with a Gaussian prior on regression weights is ridge regression. Then, for the model $y = wx + \varepsilon$ with $\varepsilon \sim N(0, 1)$, prior $w \sim N(0, 1)$ and data (x, y) = (1, 2), (2, 3), compare the MLE and MAP. (4)	**3**
5.	Differentiate between Hard margin and soft margin SVM.	**3**
6.	a) Show that XOR cannot be computed by one perceptron, then build a two-layer network of threshold neurons that computes it. (3)	**3**
7.	Five independent classifiers are each 60% accurate. Find the majority-vote accuracy, and say what changes if all five make identical predictions.	**3**
8.	a) Three users rate four movies: u1 = (5, 3, 0, 1), u2 = (4, 0, 0, 1) and u3 = (1, 1, 5, 4), where 0 means "not rated". (a) Compute the Euclidean, Manhattan and cosine measures between u1 and each of u2 and u3. (b) Compute the Jaccard similarity of the sets of rated movies. (c) Show what happens to each measure when u1 is doubled. (3) (Figure: subscripts inferred)	**3**

PART B Answer any one full question from each module. Each question carries 9 marks.

Module 1

9a.	Discuss any 4 examples of Machine learning Applications.	**5**
9b.	To fit a quadratic curve to one input x, give the feature vector at x = 3. Is the model still linear regression?	**4**

OR

10.	b) Derive the normal equation for linear regression. Then solve it for the data (1, 1), (2, 3), (3, 2), and compare with one vectorised gradient descent step from $\theta = 0$ with $\alpha = 0.1$ (step size). (4)	**9**

Module 2

11a.	c) A fraud-detection logistic regression is tested on 10,000 transactions, of which 100 are fraud. (a) A trivial model predicts "not fraud" always. (b) The trained model catches 70 frauds but also flags 140 genuine transactions. Compare the two using accuracy, precision, recall and F1, and explain what happens when the threshold is lowered. (4)	**5**
11b.	a) A node has 4 positive and 4 negative examples. A split gives children (3+, 1−) and (1+, 3−). Find the information gain. (2)	**4**

OR

Not affiliated with APJ Abdul Kalam Technological University

12a.	Consider a spam classification ML model. P(spam) = 0.3. For two words, P(w1 | spam) = 0.5, P(w2 | spam) = 0.4, P(w1 | ham) = 0.1, P(w2 | ham) = 0.2. Find P(spam | w1, w2).	**5**
12b.	Describe the different regression models and formulate any one error measurements used in regression analysis.	**4**

Module 3

13a.	Describe the working of a multilayer feed forward network	**5**
13b.	What is "Kernel trick"? Explain different types of kernel functions.	**4**

OR

14.	b) A digit-recognition network takes 28 × 28 = 784 pixel inputs, has hidden layers of 128 and 64 ReLU neurons, and has 10 softmax outputs. Count the parameters, estimate memory and computation, and compare with a single-layer softmax model. State the loss and the output rule. (3)	**9**

Module 4

15a.	Explain the different types of linkages in clustering.	**5**
15b.	b) Four points A, B, C, D have the distances AB = 2, AC = 6, AD = 10, BC = 5, BD = 9, CD = 4. Build the dendrograms by single linkage and by average linkage, giving every merge height. Compare the results. (3)	**4**

OR

16a.	The covariance eigenvalues are 4 and 1. Find the share of variance explained by the first principal component. Then, for eigenvalues 5, 3, 1, 0.5, 0.5, find how many components keep at least 95% of the variance.	**5**
16b.	a) In 5-fold cross-validation on 100 examples, how many models are trained, and how large is each training set? (2)	**4**

Not affiliated with APJ Abdul Kalam Technological University