KTUPredict practice paper, Set 2: built from the next 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. Define machine learning with suitable examples. What makes machine learning different from traditional programming. **3**

2. Define supervised learning. Name special cases of supervised learning depending on whether the inputs/outputs are categorical, or continuous. **3**

3. What is the need of validation data while training a model? **3**

4. A model has 2% training error and 25% validation error. Diagnose the problem and give two remedies. **3**

5. For a hyperplane $w^Tx + b = 0$ with $w = (3, 4)$, find the width of the margin and the distance of the point $(2, 2)$ from the hyperplane when $b = -5$. **3**

6. Choose weights and a bias for a perceptron that implements the AND gate, and verify one case that gives 1 and one that gives 0. **3**

7. Illustrate the strength and weakness of k-means algorithm. **3**

8. Describe boosting. What is the relation between Boosting and Ensemble learning? **3**

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

Module 1

9. a) Observations x1, …, xn come from Uniform(0, θ), and the largest value is 7. Find the MLE of θ, and say whether it is biased. (2) (Figure: subscripts/θ not legible in print, inferred) **9**

OR

10. c) A company has m = 10? examples. Compare gradient descent and the normal equation for (a) n = 100 features and (b) n = 10? features, using 1000 gradient descent iterations. Recommend a method for each case. (3) (Figure: exponents lost in print: m = 10^? and n = 10^? illegible; likely powers of ten such as 10^6 and 10^5) **9**

Module 2

11. c) A screening model on 10,100 samples has TPR = 0.8 and FPR = 0.05. Compare the precision when the data set has (a) 100 positives and 10,000 negatives, and (b) 5,050 of each class. Explain why ROC curves can look good while precision is poor. (4) **9**

OR

12. b) A spam filter uses MAP classification with Laplace (add-one) smoothing. The spam class has 100 words in total and the word "free" appears 9 times. The legitimate class has 200 words in total and "free" appears 2 times. The vocabulary has 50 words. Priors are P(spam) = 0.4 and P(ham) = 0.6. Classify a message containing the word "free" once. Explain why smoothing is needed. (3) **9**

Module 3

Not affiliated with APJ Abdul Kalam Technological University

13. a) The 1-D data have class +1 at x = −3 and x = 3, and class −1 at x = −1, 0 and 1. Can a single threshold separate them? Give a feature map that can. (2) **9**

OR

14. What is the gradient descent method? How is it used in the backpropagation algorithm? **9**

Module 4

15. Define Bias-Variance decomposition. **9**

OR

16. Given the following distance matrix, construct the dendrogram using agglomerative clustering with complete linkage. (Figure: Distance matrix A-E: A: 0,1,2,2,3; B: 1,0,2,4,3; C: 2,2,0,1,5; D: 2,4,1,0,3; E: 3,3,5,3,0) **9**

Not affiliated with APJ Abdul Kalam Technological University