MSc Data Science

Data Mining and Machine Learning

Final Examination, II Semester, 2019–2020 (Online)

Date : 28 August, 2020 Marks : 40
Duration : 4.5 hours, including upload of answers Weightage : 40%

1. The k-nearest neighbours (k-NN) classifier works as follows. Assume that k is odd and classification is binary. For each new input, identify the k nearest training data points and use the majority class among these to label the input. Explain whether the k-NN classifier exhibits high bias and/or variance. *(4 marks)*

2. Explain why the perceptron learning algorithm allows the use of kernel functions. *(4 marks)*

3. Show that if K_1 and K_2 are kernels, then so is $K_1 + K_2$. *(4 marks)*

4. When we use bagging, we need not keep aside a separate test set to validate our model. Explain. *(4 marks)*

5. Consider a neural network with four input features x_1, x_2, x_3, x_4 and a single output y. How many parameters do you have to estimate in the following situations, where each pair of adjacent layers is completely connected and there is a single output layer.

   (a) A shallow network with 1 hidden layer with 15 nodes.

   (b) A deep network with 5 hidden layers, with 3 nodes per layer.

   (4 marks)

6. We made the following assumptions about the loss (cost) function C for neural networks.

   - For each input x, $C(x)$ is a function of only the output layer activation.
   - The total cost across the training set is the average of the individual input costs

   Explain why these assumptions are important for effective learning of the parameters.

   (4 marks)

7. Given two input features x_1, x_2, explain how to construct a neural network to approximate a "rectangular box" function $g(x)$ with height h for $\ell_1 \leq x_1 \leq r_1$ and $\ell_2 \leq x_2 \leq r_2$. In other words, the function to be approximated is the following.

$$g(x_1, x_2) = \begin{cases} h & \text{if } \ell_i \leq x_i \leq r_i, i \in \{1, 2\} \\ 0 & \text{otherwise} \end{cases}$$

   (4 marks)

8. How can an inverted index on individual words be adapted to answer queries based on phrases (sequences of words)? *(4 marks)*

9. Explain using examples why Markov chains that are (a) not irreducible and (b) not aperiodic can fail to have a stationary distribution. *(4 marks)*

10. Consider the iterative algorithm to compute singular vectors discussed in the class. Explain why the singular values computed by the algorithm satisfy $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. *(4 marks)*