

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)
ORGANISATION OF ISLAMIC COOPERATION (OIC)
Department of Computer Science and Engineering (CSE)

SEMESTER FINAL EXAMINATION
DURATION: 1 Hour 30 Minutes

WINTER SEMESTER, 2020-2021
FULL MARKS: 75

CSE 4553: Machine Learning

Programmable calculators are not allowed. There are **3 (Three)** questions. Answer all **3** of them.

Figures in the right margin indicate marks of each question.

The square brackets on the start of each question denotes the corresponding CO and PO.

1. a) [CO2, PO3] Figure 1, shows the results of clustering the same data with k means, with k running from 2 to 6; also a plot of the sum-of-squares versus k. 4 + 3

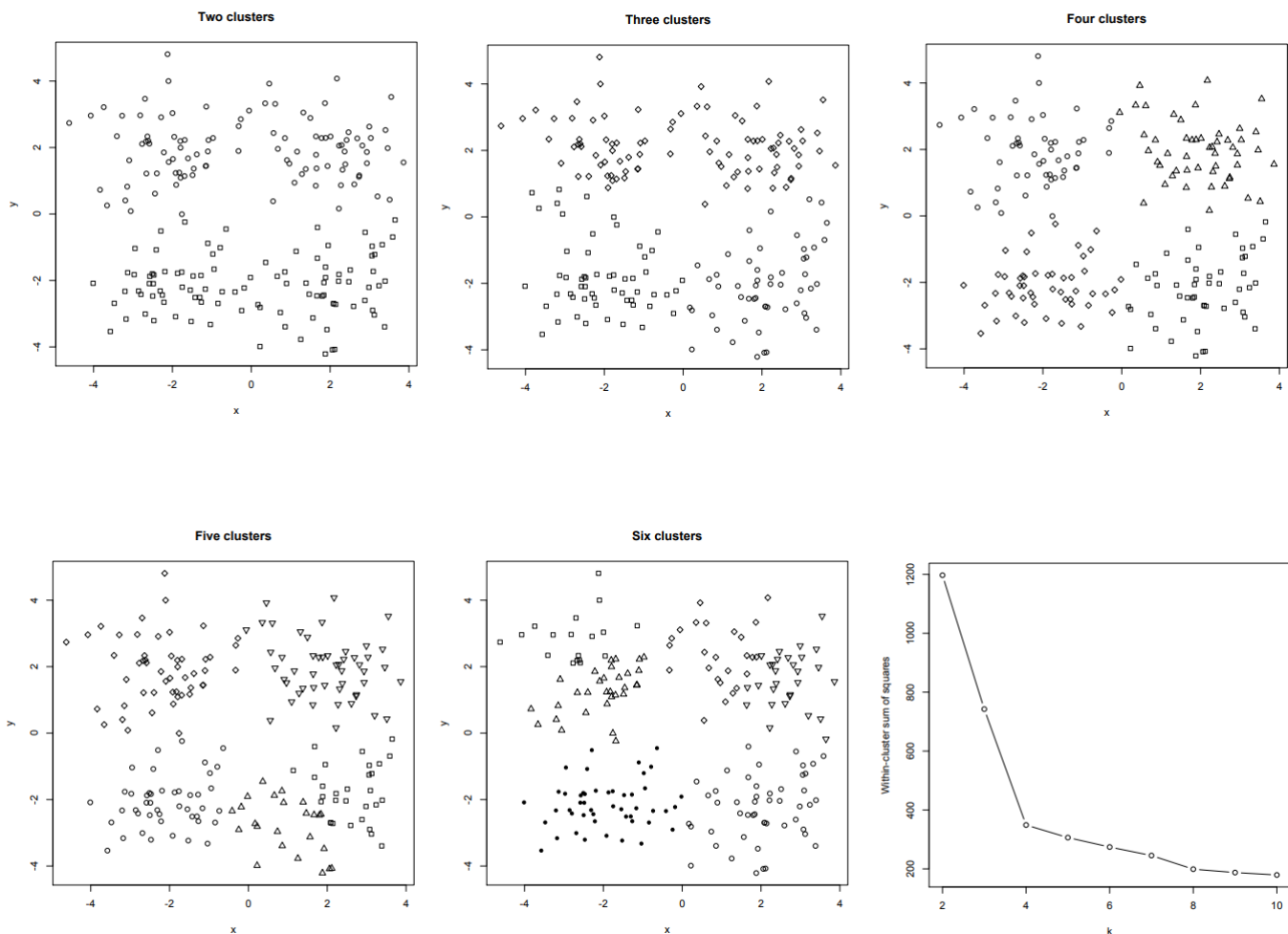


Figure 1: Clustering result for Question 1

- i. How many clusters would you guess this data has, and why?
 - ii. Does it matter whether the plot is an average over many runs of the algorithm?
- b) [CO3, PO3] Draw some points in two-dimensional space which visually seems to have two clusters, but in which K-means clustering with K=2 would be unlikely to arrive at a good solution. Explain why K-means would fail for the data you drew. 5 + 5
- c) [CO2, PO2] What drawbacks of K-means clustering are solved by K-medoids and Hierarchical clustering and how? 4 + 4

2. a) [CO2, PO2] What is overfitting? For each of the following, determine if the change would result in overfitting becoming more or less likely. 2 + 4
- An increase in the size of the dataset.
 - An increase in the complexity of the model.
 - An increase in noise in the data labels.
 - An increase in complexity in the target function.
- b) [CO1, PO1] Explain over-fitting and under-fitting issues in deep learning model training. How does the dropout layer mitigate over-fitting? 4 + 5
- c) [CO1, PO1] Write short notes on the following model training hyper-parameters. 2 × 2
- Batch Size
 - Learning Rate
- d) [CO2, PO2] Define **Precision** and **Recall**. Give one scenario where precision is preferred over recall and another scenario demanding the opposite. 2 + 4
3. a) [CO3, PO4] Figure 2 depicts training and validation curves of a learner with increasing model complexity. 3 × 2

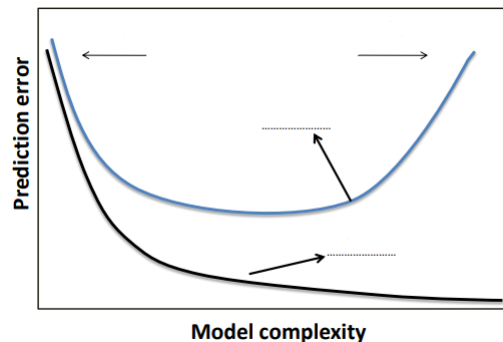


Figure 2: Model complexity vs Prediction error graph

- Which of the curves is more likely to be the training error and which is more likely to be the validation error? Indicate on the graph by filling the dotted lines.
 - In which regions of the graph are bias and variance low and high? Indicate clearly on the graph with four labels: “low variance”, “high variance”, “low bias”, “high bias”.
 - In which regions does the model overfit or underfit? Indicate clearly on the graph by labeling “overfit” and “underfit”.
- b) [CO2, PO2] Assume you have a model with a high bias and a low variance. What are the characteristics of such a model? What are the possible solutions for this problem? 2 + 4
- c) Assume that we have used a classifier for classifying documents with respect to sentiment. Classes are Pos (positive), Neg (negative) and Neu (neutral). We test our classifier on 10 documents for which gold classes (Ground truth) are given. The testing result is shown in Figure 3:
- [CO2, PO3] Draw the confusion matrix for given data. 3
 - [CO2, PO3] Compute precision, recall and accuracy for these classification results for all three classes. 6
- d) [CO1, PO1] How does Area Under the Curve (AUC) evaluate the performance of a machine learning model? 4

<i>Documents</i>	<i>gold class</i>	<i>system class</i>
d_1	<i>Pos</i>	<i>Pos</i>
d_2	<i>Pos</i>	<i>Pos</i>
d_3	<i>Pos</i>	<i>Pos</i>
d_4	<i>Pos</i>	<i>Neu</i>
d_5	<i>Neg</i>	<i>Neg</i>
d_6	<i>Neg</i>	<i>Neu</i>
d_7	<i>Neg</i>	<i>Neg</i>
d_8	<i>Neu</i>	<i>Pos</i>
d_9	<i>Neu</i>	<i>Neu</i>
d_{10}	<i>Neu</i>	<i>Neu</i>

Figure 3: Prediction result

Bonus Question - 1: Writing interesting machine learning questions is difficult. Write your own question about material covered 10-601. You will get maximum credit for writing an interesting and insightful question. [2 marks]

Bonus Question - 2: Attempt to answer your question. You will get maximum credit for providing an insightful (and correct!) answer. [2 marks]