**For Questions 1- 4, please submit a word file or a PDF file;**

**For Question 5 (programming question), please submit an .ipynb file.**

**Question 1**: **[4 points] Explain what is the bias-variance trade-off? Describe few techniques to reduce bias and variance respectively.**

Bias-Variance trade-off is the trade-off in complexity of the model. The conflicting relationship between underfitting (bias) and overfitting (variance) when the model needs to minimize bias and variance in order to find the best solution is what creates this issue. Increasing the bias will decrease the variance and increasing the variance will decrease the bias; A low complexity model will have high bias and low variance while a high complexity model will have low bias and high variance. Therefore, Bias and variance have an inverse relationship.

Diagram

Description automatically generated

Low complexity models are prone to underfitting (Accurate but initially incorrect predictions, low error) , high complexity models are prone to overfitting (Accurate, large error).

Techniques to reduce bias error;

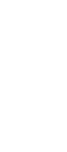
1. Change model methods (reduce high bias)
2. Train the data using multiple models (reduce high variance)
3. Feed the model more data (reduce high variance)
4. Ensemble learning methods
5. Using diverse training data

**Question 2**: [6 points] Assume the following confusion matrix of a classifier. Please compute its 1) precision,

2) recall, and 3) F1-score.

Predicted results

|  |  |  |
| --- | --- | --- |
|  | Class 1 | Class 2 |
| Class 1 | 50 | 30 |
| Class 2 | 40 | 60 |



Actual values

A picture containing text

Description automatically generated

**Question 3:** [10 points] Build a decision tree using the following training instances (using information gain approach):

Table

Description automatically generated

Graphical user interface, text

Description automatically generated with medium confidence

A picture containing line chart

Description automatically generated

**Question 4**. [10 points] The naïve Bayes method is an ensemble method as we learned in Module 5. Assuming we have 3 classifiers, and their predicted results are given in the table 1. The confusion matrix of each classifier is given in table 2. Please give the final decision using the Naïve Bayes method:

Table 1 Predicted results of each classifier

|  |  |
| --- | --- |
| Sample x | Result |
| Classifier 1 | Class 1 |
| Classifier 2 | Class 1 |
| Classifier 3 | Class 2 |

Table 2 Confusion matrix of each classifier

i) Classifier 1 ii) Classifier 2 iii) Classifier 3

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Class1 | Class2 |  |  | Class1 | Class2 |  |  | Class1 | Class2 |
| Class1 | 40 | 10 | Class1 | 2  0 | 3  0 | Class1 | 5  0 | 0 |
| Class2 | 30 | 20 | Class2 | 2  0 | 3  0 | Class2 | 4  0 | 1  0 |

A picture containing text, receipt, document

Description automatically generated