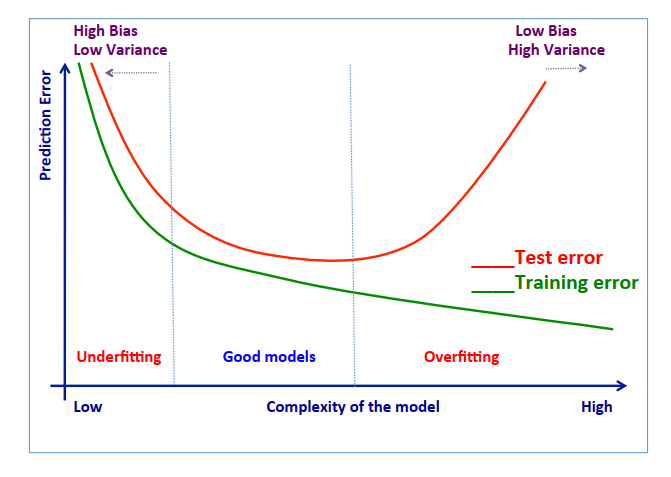
**AI Final Exam**

**Exam Date: 25-Feb-2024 9:00am**

**Submitted Date: 25-Feb-2024 12:00pm**

Theory (10pts)

1. Explain in detail what is Training Set, Validation Set, and Testing Set?
2. Regarding the figure below, what is Underfitting, Good Models, Overfitting?
3. In case of Overfitting existing, how to fix it and please explain in details?
4. Below Confusion Matrix, please calculate the value of these term and explain:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Actual | |
|  |  | Orange | Banana |
| Prediction | Orange | 45 | 5 |
| Banana | 20 | 60 |

* 1. Accuracy
  2. Precision
  3. Recall
  4. specificity
  5. FI measure

**Answer**

1.

- Training set is sample of data used to fit the model for training.

- Validation Dataset is sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration.

- Testing set is sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.

2. Underfitting, Good Models, and Overfitting: The x-axis represents the model complexity, which can be influenced by factors like the number of features used or the number of layers in a neural network. The y-axis shows the prediction error, measured by metrics like mean squared error or classification accuracy.

* Underfitting: This occurs when the model is too simple. It cannot capture the underlying patterns in the data, resulting in high training and test errors (represented by the leftmost part of the curve).

* Good Models: These models achieve a balance between complexity and accuracy. They fit the training data well without overfitting to noise or irrelevant patterns. This sweet spot is represented by the lowest point on the curve.

* Overfitting: This occurs when the model is too complex. It memorizes the training data too well, including noise and irrelevant details.

3.

4. Calculation value.

a. Accuracy = (TP +TN) / (TP + TN + FP + FN)

= (45 + 60) / (45 + 60 + 5 + 20) = 105/130 = 0.8

b. Precision = TP / (TP + FP)

For orange = 45 / (45 + 20) = 0.69

For banana = 60 / (60 + 5) = 0.92

c. Recall = TP / (TP + FN)

For orange = 45 / (45 + 5) = 0.9

For banana = 60 / (60 + 20) = 0.75

d. Specification = TN / (TN + FP)

For orange = 60 / (60 + 20) = 0.75

For banana = 45 / (45 + 5) = 0.9