**NHEAN SEREYVOUTEY**

**Final Exam**

**Answer**

1. Explain in detail

* Training Set is the set of the data on which the actual training takes place.
* Validation split helps to improve the model performance by fine-tuning the model after each epoch.
* Test set informs us about the final accuracy of the model after completing the training phase.

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.

4.

Accuracy: (TP + TN) / (TP + TN + FP + FN)

Accuracy = (45 + 60) / (45 + 20 + 5 + 60) = 0.807

Precision: TP / (TP + FP)

Precision = 60 / (60 + 20) = 0.75

Recall (Sensitivity): TP / (TP + FN)

Recall = 60 / (60+ 5) = 0.9

F1 Score: 2 \* (Precision \* Recall) / (Precision + Recall)

F1 Score = 2 \* (0.75\*0.9) / (0.75 + 0.9) = 0.81

False Positive Rate (FPR): FP / (FP + TN)

FPR = 20 / (20 + 45) = 0.3

True Positive Rate (TPR) or Recall: TP / (TP + FN)

TPR = 60 / (60 + 5) = 0.9

ROC-AUC Score: Area under the Receiver Operating Characteristic curve

ROC-AUC score needs the prediction probabilities, not just the binary predictions, to be calculated.