Math 342W/642/742W

Recitation – Day #22 (5.6.25)

I. Asymmetric Cost Binary Classification

(i) What is the typical output space of binary classification models?

(ii) What are the two erroneous decisions/classifications that can be made?

(iii) What are the two correct decisions/classifications that can be made?

(iv) Write out the meanings of each of the following abbreviations:

- N =
- P =
- PN =
- PP =
- FP =
- FN =
- TP =
- TN =
- c_{FP} =
- c_{FN} =

(v) What is the hyperparameter we are considering in the analysis of a binary classifier?

(v) Create the 2×2 Confusion Table/Matrix below:

II. Binary Classification Performance Measures

Define the following binary classification performance measures:

(i) Error:

(ii) Precision:

(iii) Accuracy:

(iv) Recall:

(v) F1 Score:

(vi) False Discovery Rate:

(vii) False Omission Rate:

(viii) Total Cost:

(ix) Specificity:

(x) False Positive Rate:

III. The ROC Curve

(i) What is the *receiver operating characteristics* (ROC) curve?

(ii) How is the *area under the curve* (AUC) related to the ROC curve?

(iii) What does a diagonal line of a ROC curve represent?

(iv) What is the ideal standard for a classifier using the ROC curve? How do we compare different classifiers knowing their ROC curves?