EVALUATION MEASURES

__/__/____

(24) Bootstrapping and Cross Validation

Bootstrapping: resampling technique used to estimate stats (like mean, variance, accuracy) or build models when you have limited data.

We repeatedly sample {with replacement} from the dataset to create multiple bootstrap samples.

We train a model on each Di → we get model M_1, M_2, M_3..
Finally we aggregate them.

→ Out of Bag samples: On avg, 63.2% of the original data appear in each bootstrap sample.
- That means roughly 36.8% are not used - these are called out of bag (OOB) samples.
- Used to estimate model accuracy

Why Bootstrap?
1. Reduces variance
2. Works well with small dataset.
3. Basis for Bagging and Boosting.

Output: Many models (ensemble)

Cross-Validation:
Estimate how well your model generalizes on unseen data by training and testing on different subsets.

Why Cross Validation?
1. Reliable estimate of true performance
2. Helps detect overfitting
3. For model selection and hyperparameter tuning.

__/__/____

(25) 2- Class Evaluation Metrics

1. Confusion Matrix

	Pred Positive	Pred Negative
Actual Positive	True Pos	False Neg
Actual Negative	False Pos	True Neg

FP = Type I Error | FN = Type II error (Issue)

2. Accuracy : Measures overall fraction of correct prediction
- Can be misleading if classes are imbalanced

$$\text{Accuray} = \frac{TP + TN}{TP + TN + FP + FN}$$

3. Precision: Of all predicted positives, how many are correct?

$$\text{Precision} = TP / TP + FP$$

Medical Diagnosis

3. Recall (Sensitivity) : Of all actual positives, how many did we correctly predict

$$\text{Recall} = TP / TP + FN$$

Fraud Detection

4. Specificity : Of actual negatives, how many did we correctly predict as negative

$$\text{Specificity} = FN / TN + FP$$

5. F1- Score : Harmonic mean of precision and recall
- Good for imbalance datasets

$$F1 = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

__/__/____

(26) ROC and AUC

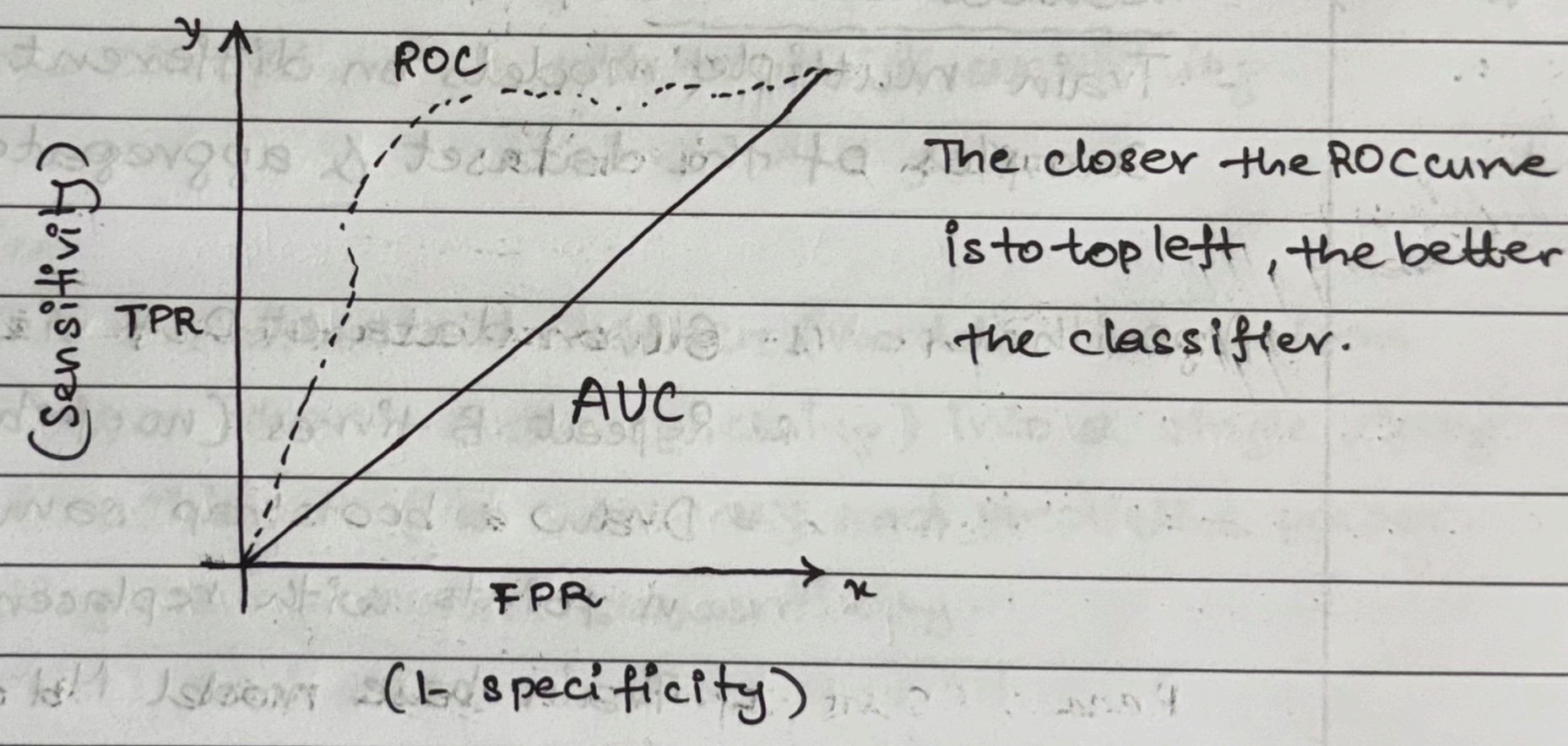
ROC Curve - Receiver Operating Characteristic Curve

+ Evaluates a classifiers performance across all thresholds rather than a single threshold.

TPR = TP / TP+FN (Sensitivity against FPR)

FPR = FP / FP+TN

Threshold: The cutoff probability at which we classify as +ve



AUC Curve: A single-number metric that summarizes ROC Curve.

- AUC = 1 → perfect classifier
- AUC = 0.5 → random classifier
- AUC = 0.5 < AUC < 1 → Better than random

Probability that a randomly chosen positive sample is ranked higher than a randomly chosen negative sample.

→ Precision-Recall curve works well when classes are imbalanced.

→ We require both ROC and AUC to give statement on AUC

AUC = 0.5 Random Model ? ROC curve can flip labels