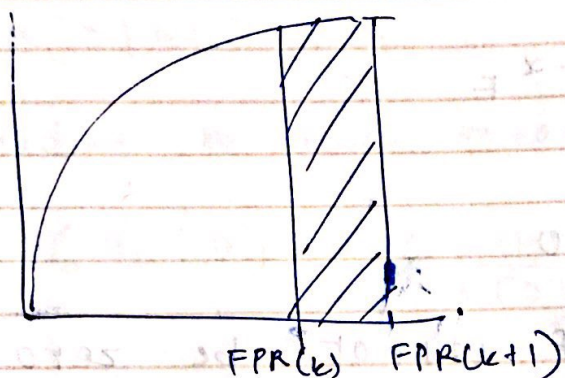


Problem 5

AUC-ROC \approx proportion of correct pairs = not inversion



$$\begin{aligned} \text{TPR} &= \frac{\#(y=1) \text{ above threshold}}{\#(y=1) \text{ total}} \\ &= \frac{\sum_{i=k}^n [y_{(i)} = 1]}{n_+} \end{aligned}$$

$$\text{FPR} = \frac{\#(y=0) \text{ above threshold}}{\#(y=0) \text{ total}} = \frac{\sum_{i=k}^n [y_{(i)} = -1]}{n_-}$$

$$\text{Area} = \frac{(FPR(k+1) - FPR(k)) \cdot (TPR(k) + TPR(k+1))}{2}$$

ROC = (FPR(t), TPR(t)) where t = threshold.

n_+ = number of samples in +ve class

n_- = number of samples in -ve class

$$AUC = \sum_{k=1}^{n-1} \frac{TPR_{k+1} + TPR_k}{2} (FPR_k - FPR_{k+1})$$

$$= \sum_{k=1}^{n-1} \frac{\sum_{i=k+1}^n [y_{(i)} = +1] + \frac{1}{2} [y_{(k)} = 1]}{n_+} \frac{[y_{(k)} = -1]}{n_-}$$

$$= \frac{1}{n_+ n_-} \sum_{k=1}^{n-1} \sum_{i=k+1}^n [y_{(i)} = 1] [y_{(k)} = -1]$$

$$= \frac{1}{n_+ n_-} \sum_{k < i} [y_{(k)} < y_{(i)}]$$

AUC of classifier is probability that the classifier will rank a randomly chosen positive example higher than a randomly chosen -ve example.

$$P(\text{score}(+) > \text{score}(-))$$