MANIPAL SCHOOL OF INFORMATION SCIENCES
MANIPAL
(A constituent unit of MAHE, Manipal)

AML5201 | Advanced Applications of Probability & Statistics | Problem Set-1

231057015 Aditya N Bhatt

1. In a future society, a machine is used to predict a crime before it occurs. If you were responsible for tuning this machine, what evaluation metric would you want to maximize to ensure no innocent people (people not about to commit a crime) are imprisoned?

➔ **Precision**

2. Consider a classification model that separates email into two categories: "spam" or "not spam." Answer the following questions regarding precision and recall (a.k.a. sensitivity or true positive rate) by playing around with the threshold slider on the demo website here:

   (a) Which is a more relevant performance metric in this case: recall or precision? Explain briefly why.

➔ **Precision**

(b) Increasing the classification threshold generally $\underbrace{\text{increases/decreases}}_{\text{choose one}}$ FP.

➔ **Decreases**

(c) When the classification threshold increases, precision

$$\underbrace{\text{probably increases/probably decreases/definitely increases/definitely decreases}}_{\text{choose one}}.$$

➔ **Probably increases.**

(d) Keeping in mind that $TP + FP + TN + FN = n$, which is the number of samples, when the classification threshold is increased, what happens to the quantity TP?

➔ **When the classification threshold is increased, the quantity TP should probably decrease.**

(e) When the classification threshold is increased, the quantities TN and FN both

$$\underbrace{\text{uniformly/non-uniformly}}_{\text{choose one}} \ \underbrace{\text{increase/decrease}}_{\text{choose one}}.$$

➔ **Non-uniformly Increase.**

(f) Decreasing the classification threshold generally $\underbrace{\text{increases/decreases}}_{\text{choose one}}$ FN.

➔ **Increase**

(g) When the classification threshold is decreased, recall

$$\underbrace{\text{probably increases/probably decreases/definitely increases/definitely decreases}}_{\text{choose one}}.$$

➔ **Probably Increases**

(h) When the classification threshold is decreased, the quantities TP and FP both

$$\underbrace{\text{uniformly/non-uniformly}}_{\text{choose one}} \ \underbrace{\text{increase/decrease}}_{\text{choose one}}.$$

➔ **Non-uniformly Increase.**

3. In which of the following scenarios would a high accuracy value suggest that the ML model is doing a good job? Explain your answer briefly.

(a) An expensive and critical hydro-electric turbine operates 23 hours a day. An ML model evaluates vibration patterns and predicts when the turbine is operating without anomaly with an accuracy 99.99%.

➔ A high accuracy value suggests the ML model is doing well as it correctly predicts the turbine's operation most of the time.

(b) You are building an ML tool for a retail company which will predict, based on past purchase history and other demographic information, the high end cellphone that the next buyer will potentially buy from an available 10 high end models. Your ML model has an accuracy of 15%.

➔ accuracy of 15% is quite low for an ML model predicting which high-end cell phone a buyer will purchase, suggesting the model isn't performing well.

(c) A deadly, but curable, medical condition afflicts .01% of the population. Your ML model uses symptoms as features and predicts this affliction with an accuracy of 99.99%.

➔ Despite the high accuracy of 99.99%, the model may not be effective due to the rare occurrence (0.01%) of the condition.

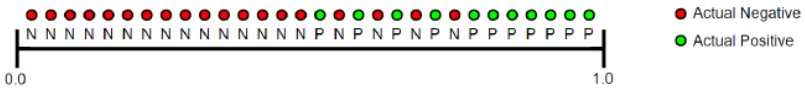
4. Consider two models: A and B, that each evaluate the same dataset. Which one of the following statements is true?
   (a) If model A has better precision and better recall than model B, then model A is probably better.
   (b) If model A has better recall than model B, then model A is better.
   (c) If Model A has better precision than model B, then model A is better.

➔ (A) is true.

5. An ROC curve is a plot *of* ***True Positive Rate (TPR)*** vs. ***False Positive Rate (FPR)*** for different ***classification threshold.***

6. Lowering the classification threshold classifies more items as positive/negative, thus increasing both ***True Positives*** and ***False Positives***.

7. AUC (Area under the ROC Curve) provides an aggregate measure of performance across all possible classification thresholds. One way of interpreting AUC is as the probability that the model ranks a random positive example more highly than a random negative example. For example, given the following examples, which are arranged from left to right in ascending order of prediction probabilities:



AUC represents the probability that a random positive (green) example is positioned to the right of a random negative (red) example. AUC ranges in value from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0.0; one whose predictions are 100% correct has an AUC of 1.0.

Suppose we multiplied all of the predictions from a given model by 0.5 (for example, if the model predicts 0.4, we multiply by 0.5 to get a prediction of 0.2), how would it change the model's performance as measured by AUC?

(a) It would make AUC terrible, since the prediction values are now way off.

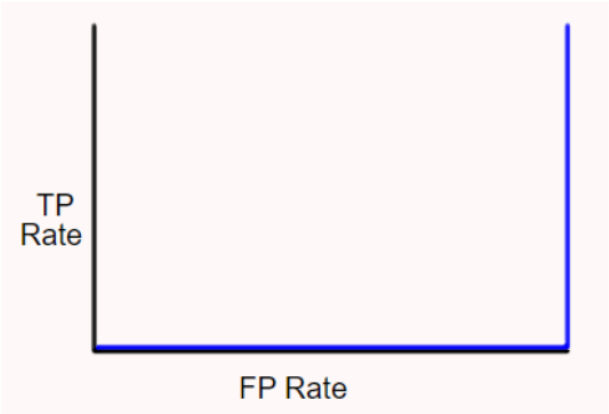
(b) It would make AUC better, because the prediction values are all farther apart.

(c) No change. AUC only cares about relative prediction probabilities.

➔ (c) No change. AUC only cares about relative prediction probabilities.

- This is because AUC measures the ability of the model to rank positive examples higher than negative ones, regardless of the actual prediction values.

8. Your friend shows you his model's ROC curve as follows:



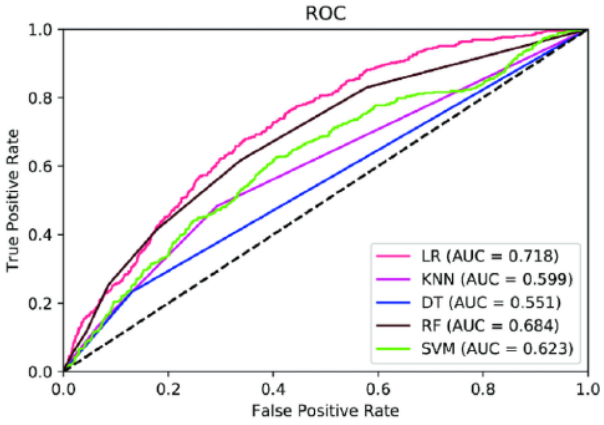
Is your friend's model any good, why? How can you help your friends model go from zero to hero?

➔ No, the Model is not performing well.

➔ To improve, consider

- more data,
- feature engineering,
- trying different models,
- tuning hyperparameters,
- using ensemble methods.

9. The figure shows ROC curves for different models.



Classify the following statements as true or false:

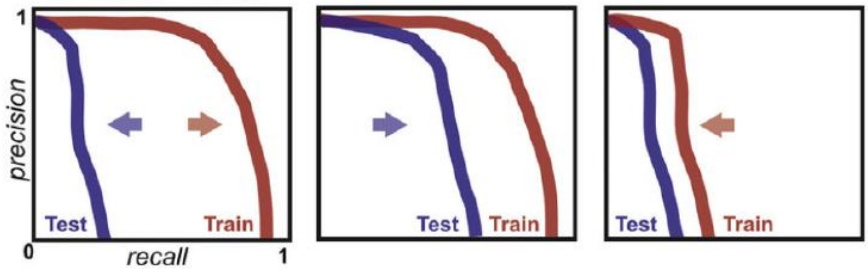
- Dashed black line represents random classification.
- ROC curve for any model can't fall below the dashed black line.
- The model represented by solid blue line is better than that represent by solid lime.

➔ (a) True
➔ (b) True
➔ (c) False

10. Which one among TP, TN, FP, FN does not play a role in forming the precision-recall curve? What does the conclusion mean intuitively?

➔ True Negatives (TN) does not play a role in forming the Precision-Recall curve.

11. The figure shows precision-recall curves for different models on the train and test sets.



Identify which model overfits, which one underfits, and which one is a good fit.

➔ The first model overfits,
➔ The second model underfits
➔ The third model is a good fit.

12. Explain which one among area under ROC and area under precision-recall curve would you use for the following scenarios:
    (a) Identifying whether a customer will buy a product on discount or not when a customer is equally likely to do so.

➔ Area Under the ROC curve (AUROC)

(b) Identifying a spam email when generally spam emails constitute 1% of the total emails.

➔ Area Under the Precision-Recall curve (AUPRC)