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

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?

➢ Let's consider **True: Person is guilty**

False: Person is innocent

Here, we are interested in bringing down the false positive rate, hence maximize the **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.

➢ **Recall** is more relevant metric. Because we are concerned with not missing out on mails and also on irrelevant mails bothering us. Using recall we can correctly classify the spam mails as spam.

b) Increasing the classification threshold generally increases/decreases FP.

➢ **Decreases**

c) When the classification threshold increases, precision:

Probably increases/probably decreases/definitely increases/definitely decreases.

➢ **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?

➢ **TP decreases when the threshold is increased.**

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

uniformly/non-uniformly increase/decrease

➢ **Non uniformly increase**

f) Decreasing the classification threshold generally increases/decreases FN:

- **Decreases**

g) When the classification threshold is decreased, recall

Probably increases/probably decreases/definitely increases/definitely decreases

- **Definitely increases.**

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

uniformly/non-uniformly increase/decrease

- **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%:

- In this case, a high accuracy would mean that the model is doing well. Because it is important that an expensive and critical machine run without any anomalies.

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 cell phone that the next buyer will potentially buy from an available 10 high end models. Your ML model has an accuracy of 15%:

- In this case, a high accuracy would suggest which high end model out of the 10 will be bought by a buyer. It is essential to know this to manage the cost of production of that particular high end cell phone.

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%.

- In this case, an accuracy of 99.99% means that the decease has afflicted 99.99% of the population, where as in reality it has only afflicted 0.01% of the population. So a high accuracy means that the model is not doing well.

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) If model A has better precision and better recall than model B; then model A is probably better.

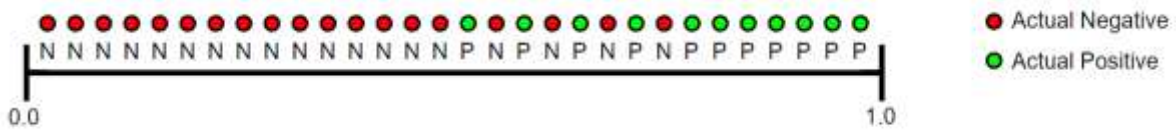
5. An ROC curve is a plot of _____ vs. _____ for different _____ :

➢ An ROC curve is a plot of **True Positive Rate** vs. **False Positive Rate** for different **classification thresholds.**

6. Lowering the classification threshold classifies more items as positive/negative thus increasing both _____ and _____:

➢ Positive, TP and FP

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:

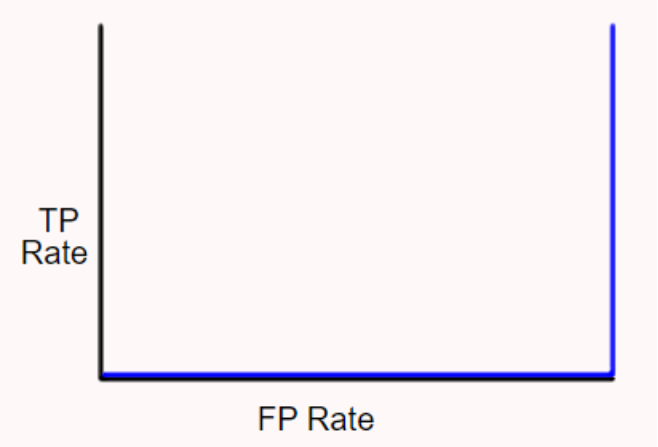


AUC represents the probability that a random positive (green) example is positioned to the right of a random negative (red) example. AUC ranges in value from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0.0; one whose predictions are 100% correct has an AUC of 1.0. Suppose we multiplied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

- It would make AUC terrible, since the prediction values are now way off.

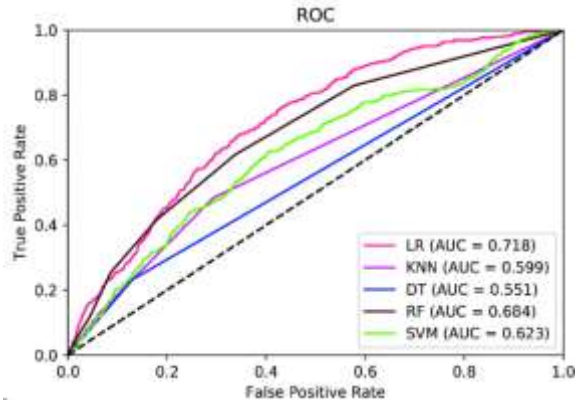
8. Your friend shows you his model's ROC curve as follows:



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

- This ROC curve is very bad because it has an AUC of 0.0. It ranks all the negatives above the positives. If we were to reverse the prediction, like flip the positives to negatives and negatives to positives the model will go from 0 to hero.!

9. The figure shows ROC curves for different models.



Classify the following statements as true or false:

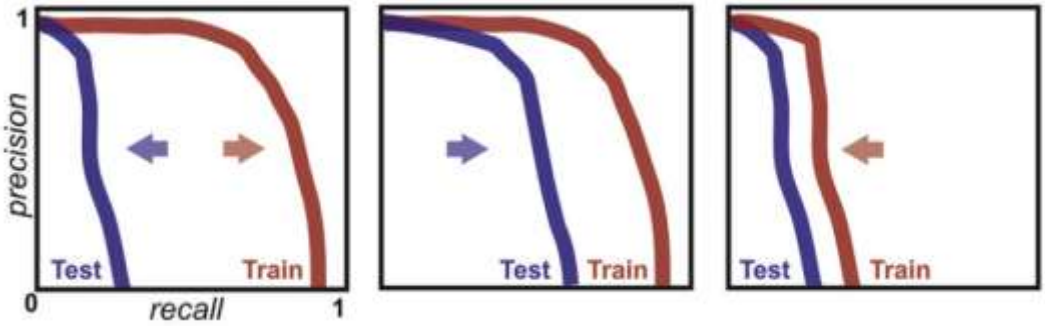
- Dashed black line represents random classification. **True**
- ROC curve for any model can't fall below the dashed black line. **True**

- The model represented by solid blue line is better than that represent by solid lime. **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?

➢ TN does not play a role forming the precision-recall curve. Intuitively this means that we are not concerned about the model predicting the correctly classifying the negative samples as negative.

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.

➢ Figure 1 **overfits.**

Figure 2 is a **good fit.**

Figure 3 is **underfits.**

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 ROC curve.

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

➢ Precision-Recall curve.