CPSC 330 Lecture 9: Classification Metrics

Varada Kolhatkar

Announcements

- Important information about midterm 1
  - https://piazza.com/class/m01ukubppof625/post/249
- HW4 has been released. Due next week Monday.
- HW5 will be released next week Tuesday. It’s a project-type assignment and you get till Oct 28th to work on it.

ML workflow

What question do I want to answer?

Formulation to supervised machine learning problem

Supervised ML process

1. Data collection 

2. Data cleaning, splitting 

3. Data exploration (EDA) 

4. Preprocessing, Feature engineering 

5. Feature selection, Model building 

6. Evaluation, model selection 

7. Test data predictions, visualizations, possible deployment 

Accuracy

- So far we have been measuring model performance using **Accuracy**.
- **Accuracy** is the proportion of all classifications that were correct, whether *positive* or *negative*.

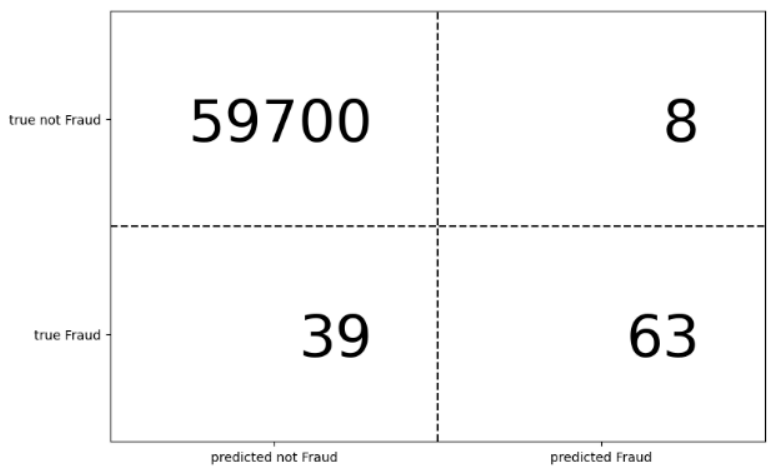
$$\text{Accuracy} = \frac{\text{corrct classifications}}{\text{total classifications}}$$

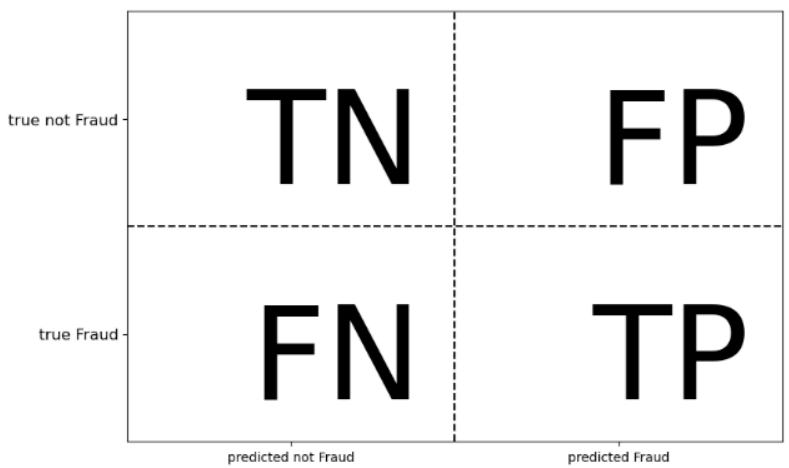
- However, in many real-world applications, the dataset is imbalanced or one kind of mistake is more costly than the other
- In such cases, it's better to optimize for one of the other metrics instead.

Fraud Confusion matrix

- Which types of errors would be most critical for the bank to address?

Fraud Confusion matrix





- TN ⟶ True negatives
- FP ⟶ False positives
- FN ⟶ False negatives
- TP ⟶ True positives

Confusion matrix questions

Imagine a spam filter model where emails classified as spam are labeled 1 and non-spam emails are labeled 0. If a spam email is incorrectly classified as non-spam, what is this error called?

- a. A false positive
- b. A true positive
- c. A false negative
- d. A true negative

Confusion matrix questions

In an intrusion detection system, intrusions are identified as 1 and non-intrusive activities as 0. If the system fails to identify an actual intrusion, wrongly categorizing it as non-intrusive, what is this type of error called?

- a. A false positive
- b. A true positive
- c. A false negative
- d. A true negative

Confusion matrix questions

In a medical test for a disease, diseased states are labeled as 1 and healthy states as 0. If a healthy patient is incorrectly diagnosed with the disease, what is this error known as?

- a. A false positive
- b. A true positive
- c. A false negative
- d. A true negative

Precision, Recall, F1-Score

TN → True Negatives (non-fraud predicted as as non-fraud)
TP → True Positives (fraud predicted as fraud)
FN → False Negatives (fraud predicted as non-fraud)
FP → False Positives (non-fraud predicted as fraud)

iClicker Exercise 9.1

iClicker cloud join link: https://join.iclicker.com/VYFJ

Select all of the following statements which are TRUE.

- a. In medical diagnosis, false positives are more damaging than false negatives (assume “positive” means the person has a disease, “negative” means they don’t).
- b. In spam classification, false positives are more damaging than false negatives (assume “positive” means the email is spam, “negative” means they it’s not).
- c. If method A gets a higher accuracy than method B, that means its precision is also higher.
- d. If method A gets a higher accuracy than method B, that means its recall is also higher.

Counter examples

Method A - higher accuracy but lower precision

Negative	Positive
90	5
5	0

Method B - lower accuracy but higher precision

Negative	Positive
80	15
0	5

Thresholding

- The above metrics assume a fixed threshold.
- We use thresholding to get the binary prediction.
- A typical threshold is 0.5.
  - A prediction of 0.90 ⟶ a high likelihood that the transaction is fraudulent and we predict **fraud**
  - A prediction of 0.20 ⟶ a low likelihood that the transaction is non-fraudulent and we predict **Non fraud**
- **What happens if the predicted score is equal to the chosen threshold?**
- Play with classification thresholds

iClicker Exercise 9.2

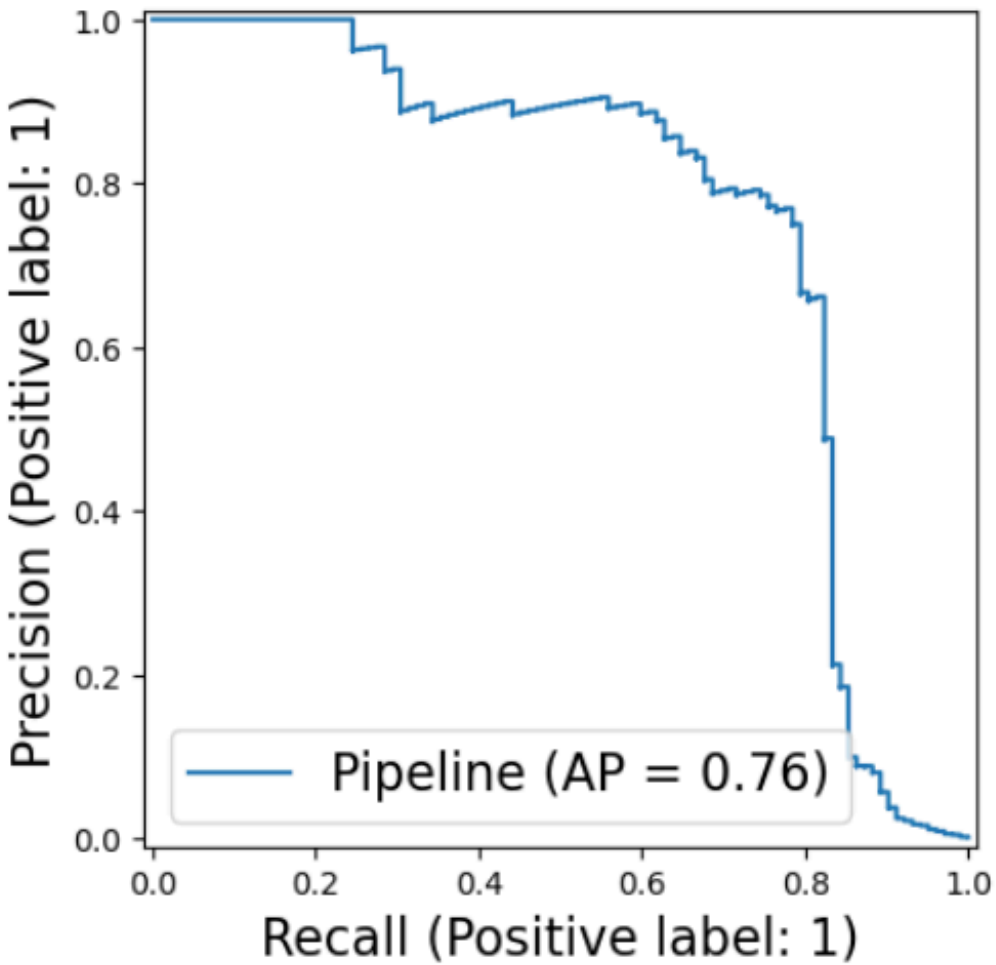
iClicker cloud join link: https://join.iclicker.com/VYFJ

Select all of the following statements which are TRUE.

- a. If we increase the classification threshold, both true and false positives are likely to decrease.
- b. If we increase the classification threshold, both true and false negatives are likely to decrease.
- c. Lowering the classification threshold generally increases the model's recall.
- d. Raising the classification threshold can improve the precision of the model if it effectively reduces the number of false positives without significantly affecting true positives.

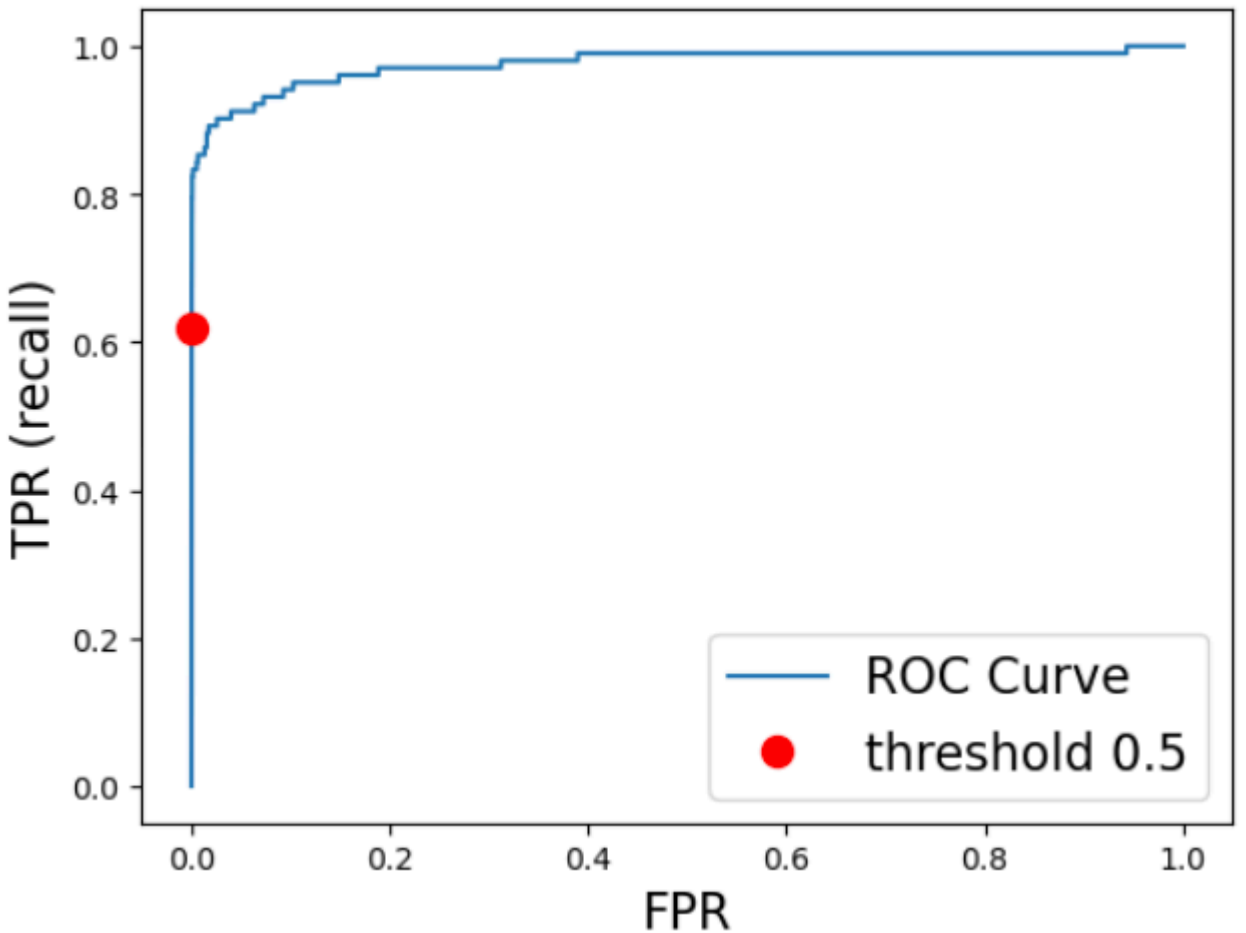
PR curve

- Calculate precision and recall (TPR) at every possible threshold and graph them.
- Better choice for highly imbalanced datasets



ROC curve

- Calculate the true positive rate (TPR) and false positive rate (FPR) at every possible thresholding and graph TPR over FPR.
- Good choice when the datasets are roughly balanced.

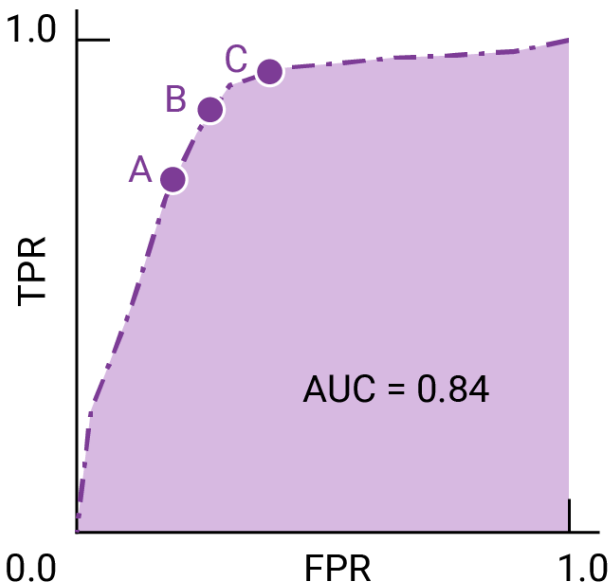


AUC

- The area under the ROC curve (AUC) represents the probability that the model, if given a randomly chosen positive and negative example, will rank the positive higher than the negative.

ROC AUC questions

Consider the points A, B, and C in the following diagram, each representing a threshold. Which threshold would you pick in each scenario?



- a. If false positives (false alarms) are highly costly
- b. If false positives are cheap and false negatives (missed true positives) highly costly
- c. If the costs are roughly equivalent

Source