Performance Metrics

CS277

Class Imbalance problems

- Lots of classification problems where the classes are skewed (more records from one class than another)
  - Credit card fraud Intrusion detection
  - Defective products in manufacturing assembly line
  - COVID-19 test results on a random sample
- Key Challenge:
  - Evaluation measures such as accuracy are not well suited for imbalanced class

Metrics for Performance Evaluation

- Focus on the predictive capability of a model
  - Rather than how fast it takes to classify or build models, scalability, etc.
- Confusion Matrix:

	PREDICTED CLASS			
ACTUAL CLASS		Class=Yes	Class=No	Total
	Class=Yes	a	b	a+b
	Class=No	c	d	c+d
	Total	a+c	b+d	

a: TP (true positive)

b: FN (false negative)

c: FP (false positive)

d: TN (true negative)

Metrics for Performance Evaluation...

	PREDICTED CLASS	
	Class=Yes	Class=No
ACTUAL CLASS: Class=Yes	a (TP)	b (FN)
ACTUAL CLASS: Class=No	c (FP)	d (TN)

Precision = $\frac{a}{a+c}$

Recall or Sensitivity = $\frac{a}{a+b}$

Specificity = $\frac{d}{c+d}$

- Most widely-used metric:

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Error Rate} = \frac{b+c}{a+b+c+d} = \frac{FP+FN}{TP+TN+FP+FN}$$

LIMITATION OF ACCURACY

- Consider a 2-class problem
  - Number of Class 0 examples = 9990
  - Number of Class 1 examples = 10

- If model predicts everything to be class 0, then what is the accuracy?
- It is 9990/10000 = 99.9 %
  - Accuracy is misleading because model does not detect any class 1 example

Which model is better?

A

	PREDICTED	
ACTUAL	Class=Yes	Class=No
Class=Yes	0	10
Class=No	0	990

Accuracy: 99%

B

	PREDICTED	
ACTUAL	Class=Yes	Class=No
Class=Yes	10	0
Class=No	500	490

Accuracy: 50%

Which model is better?

A

	PREDICTED	
	Class=Yes	Class=No
ACTUAL Class=Yes	5	5
ACTUAL Class=No	0	990

B

	PREDICTED	
	Class=Yes	Class=No
ACTUAL Class=Yes	10	0
ACTUAL Class=No	500	490

METRICS FOR PERFORMANCE EVALUATION…

	PREDICTED CLASS	
ACTUAL CLASS	Class=Yes	Class=No
Class=Yes	a (TP)	b (FN)
Class=No	c (FP)	d (TN)

Precision (p) = $\frac{a}{a+c}$

Recall (r)= $\frac{a}{a+b}$

F-measure (F) = $\frac{2rp}{r+p} = \frac{2a}{2a+b+c}$

Accuracy $= \frac{a+d}{a+b+c+d}$

Specificity = $\frac{d}{c+d}$

Recall is also known as Sensitivity

	PREDICTED CLASS	
		Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	10	0
	Class=No	10	980

Precision (p) $= \frac{a}{a+c} = \frac{10}{10+10} = \frac{1}{2} = 0.5$

Recall (r)$= \frac{a}{a+b} = \frac{10}{10+0} = 1$

F-measure (F) $= \frac{2rp}{r+p} = \frac{2a}{2a+b+c} = \frac{2*10}{2*10+0+10} = \frac{20}{30} = \frac{2}{3} = 0.67$

Accuracy $= \frac{a+d}{a+b+c+d} = \frac{10+980}{10+0+10+980} = 0.99$

	PREDICTED CLASS	
		Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	1	9
	Class=No	0	990

Precision (p) $= \frac{a}{a+c} = \frac{1}{1+0} = 1$

Recall (r)$= \frac{a}{a+b} = \frac{1}{1+9} = 0.1$

F-measure (F) $= \frac{2rp}{r+p} = \frac{2a}{2a+b+c} = \frac{2*1}{2*1+9+0} = \frac{2}{11} = 0.18$

Accuracy $= \frac{a+d}{a+b+c+d} = \frac{1+990}{1+9+0+990} = \frac{991}{1000} = 0.991$

	PREDICTED CLASS		
		Yes	No
ACTUAL CLASS	Yes	TP	FN
	No	FP	TN

α is the probability that we reject the null hypothesis when it is true. This is a Type I error or a false positive (FP).

β is the probability that we accept the null hypothesis when it is false. This is a Type II error or a false negative (FN).

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

$$ErrorRate = 1 - accuracy$$

$$Precision = Positive\ Predictive\ Value = \frac{TP}{TP + FP}$$

$$Recall = Sensitivity = TP\ Rate = \frac{TP}{TP + FN}$$

$$Specificity = TN\ Rate = \frac{TN}{TN + FP}$$

$$FP\ Rate = \alpha = \frac{FP}{TN + FP} = 1 - specificity$$

$$FN\ Rate = \beta = \frac{FN}{FN + TP} = 1 - sensitivity$$

$$Power = sensitivity = 1 - \beta$$

A	PREDICTED CLASS		
ACTUAL CLASS		Class=Yes	Class=No
	Class=Yes	40	10
	Class=No	10	40

$\text{Precision (p)} = 0.8$

$\text{TPR} = \text{Recall (r)} = 0.8$

$\text{FPR} = 0.2$

$\text{F-measure (F)} = 0.8$

$\text{Accuracy} = 0.8$

$$\frac{\text{TPR}}{\text{FPR}} = 4$$

B	PREDICTED CLASS		
ACTUAL CLASS		Class=Yes	Class=No
	Class=Yes	40	10
	Class=No	1000	4000

$\text{Precision (p)} = 0.038$

$\text{TPR} = \text{Recall (r)} = 0.8$

$\text{FPR} = 0.2$

$\text{F-measure (F)} = 0.07$

$\text{Accuracy} = 0.8$

$$\frac{\text{TPR}}{\text{FPR}} = 4$$

SOME INTERPRETATIONS

- Low recall

  low recall means a high number of false negatives

  - A low recall means a high number of FN

- High recall

  Recall : It focuses on the model's ability to identify all positive cases, even if it results in some false positives.

  - A high recall means that most of the positive cases (TP+FN) will be labeled as positive (TP).

- Low precision

  Getting correct prediction is low if precision is low

  - Very less chance that a predicted case is positive

  low precision means there is very low chance that predicted value is positive

- High precision
  - Very high chance that a predicted case is positive

AUC (area under curve)-ROC (receiver operating curve)

- ROC is a performance measurement for the classification problems at various threshold settings

helps us to find the best threshold for classification

- ROC is a probability curve

AUC is maximum means best threshold

- AUC represents the degree or measure of severability
- Developed in 1950s for signal detection theory to analyze noisy signals
- ROC curve plots TPR against FPR
  - Performance of a model represented as a point in an ROC curve

$$\text{True Positive Rate} = \frac{TP}{TP+FN}$$

Recall, Sensitivity

$$\text{Specificity} = \frac{TN}{FP+TN}$$

$$\text{False Positive Rate} = \frac{FP}{FP+TN}$$

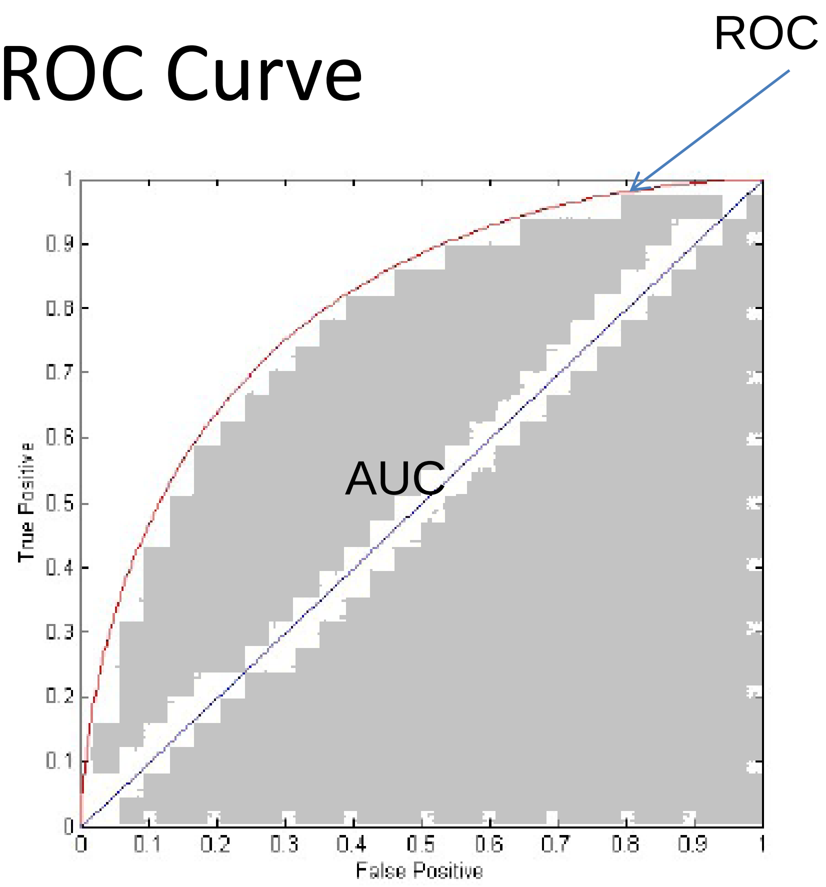
1-Specificity

AUC-ROC Curve

(TPR,FPR):

- (0,0): declare everything to be negative class
- (1,1): declare everything to be positive class
- (1,0): ideal

- Diagonal line:
  - Random guessing
  - Below diagonal line:
    - prediction is opposite of the true class



ROC Curve

- To draw ROC curve, classifier must produce continuous-valued output
  - Outputs are used to rank test records, from the most likely positive class record to the least likely positive class record
  - By using different thresholds on this value, we can create different variations of the classifier with TPR/FPR tradeoffs
- Many classifiers produce only discrete outputs (i.e., predicted class)
  - How to get continuous-valued outputs?
  - Decision trees, rule-based classifiers, neural networks, Bayesian classifiers, k-nearest neighbors, SVM