Classification Performance Metrics

Type	Lecture Notes
Reviewed	☐
Available Summary?	Done
Week	3

Metrics for classifier performance

		actual class	
		class 1 (positive)	class 2 (negative)
predicted class	class 1 (positive)	21 (TP)	6 (FP)
	class 2 (negative)	7 (FN)	41 (TN)

TN: true negatives FN: false negatives

FP: false positives TP: true positives

Accuracy

the percentage of all the observations the system labeled correctly

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

limitations:

- in a binary classification; if model predicts everything to be class 1, but model never will correctly predict any sample in class 0
- in imbalanced classification problems case accuracy is misleading and does not give a good picture of model quality

Misclassification rate

the percentage of all the observations the system labeled wrong

$Misclassification\_rate = \frac{FP+FN}{TP+TN+FP+FN}$

confusion matrix

types of errors:

False positives:

the system predicted TRUE but the value was FALSE (aka False Alarms or Type I Error)

False negatives:

the system predicted FALSE but the value was TRUE (aka Misses or Type II Error)

Precision

how many predicted labels were true

aka. positive predictive value (PPV)

$Precision = \frac{TP}{TP+FP}$

Recall

how many of the true events are identified

aka true positive rate (TPR) or sensitivity

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

F-measure

harmonic mean of precision and recall

$F - measure = \frac{2.P.R}{P+R}$

can be wighted to favor precision or recall

$F_\beta = \frac{(1+\beta).P.R}{\beta^2 P+R}$

beta> 1 favors recall

beta<1 favors precision

Matthew's correlation coefficient

MCC can be understood as a specific case of a linear correlation coefficient Pearson r for a binary classification setting and is considered as especially useful in unbalanced class settings

The previous metrics take values in the range between -1 (worst) and 1 (best) whereas the MCC is bounded between the range 1 (perfect correlation between ground truth and predicted outcome) and -1 (inverse or negative correlation) value of 0 denotes a random prediction

$$MCC = \frac{TP.TN - FP.FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Knob (threshold)

Most classifiers have a knob or threshold that you can adjust

ROC (Receiver Operating Characteristic)

- ROC curve plots TPR (on the y-axis) against FPR (on the x-axis)

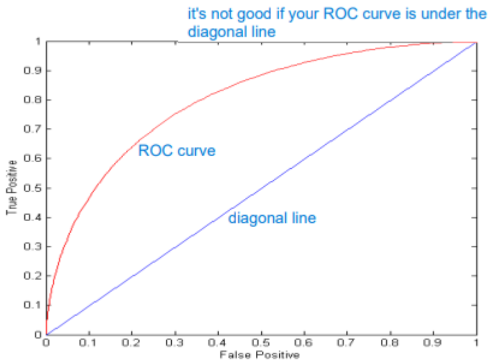
$TPR = \frac{TP}{TP+FN}$ (recall)

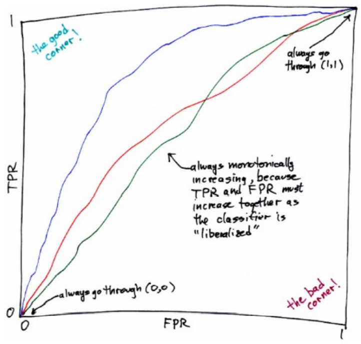
- Fraction of positive instances predicted correctly

$FPR = \frac{FP}{FP+TN}$

Fraction of negative instances predicted correctly

- Performance of a classifier represented as a point on the ROC curve
- Changing some parameter of the algorithm changes the location of the point





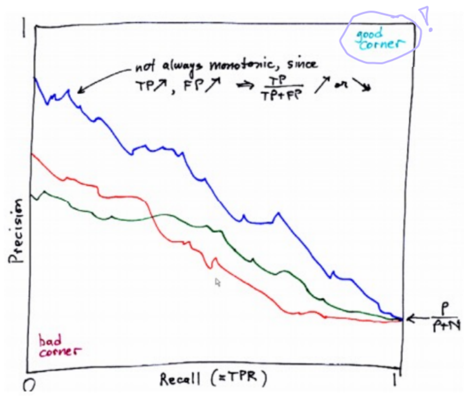
blue dominates red and green
neither red nor green dominate the other

Area under the ROC curve: AUC

Ideal: Area = 1

Random Guess: Area = 0.5

Precision Recall Curves



False discovery rate

$FDR = \frac{TP}{TP+FP}$

Negative predictive value

$NPV = \frac{TN}{FN+TN}$

Specificity

$SPC = \frac{TN}{FP+TN}$

Micro vs Macro Averaging

> *the size of the classes is different between micro and macro avg*

Macro-avg

- compute performance for each class then average

Micro-avg

- collect decisions for all classes, compute contingency table, evaluate
- dominated by score on common classes

Class 1

	Truth: yes	Truth: no
Classifier: yes	10	10
Classifier: no	10	970

Class 2

	Truth: yes	Truth: no
Classifier: yes	90	10
Classifier: no	10	890

Micro Ave. Table

	Truth: yes	Truth: no
Classifier: yes	100	20
Classifier: no	20	1860

- Macroaveraged precision: $(\overset{\text{prec of class1}}{0.5} + \overset{\text{prec of class2}}{0.9})/2 = 0.7$
- Microaveraged precision: $\underset{\text{all of the TP}}{100}/\underset{\text{all of the TP + TN}}{120} = .83$

▼ Hit @k performance metrics

consider the top K predictions and count it as TP positive if correct class is in the top K prediction