How to Define the Effectiveness a Classification Model?

Harvey Alférez, Ph.D.

Introduction

- Place objects into appropriate category based on attributes
- Object: credit card
- Two categories: legitimate, fraud (binary classification - place an object into one of two possible groups)
  - Typically in statistics, the **classes** are called ***null hypothesis*** and ***alternative hypothesis***

Type 1 vs. Type 2 Errors

- Claim is legitimate but prediction = fraud: type 1 - leads to inconvenience
- Claim is a fraud but prediction = legitimate: type 2 - leads to financial lost

Defining the Effectiveness of a Classifier

Is this transaction a fraud? (+ yes fraud, - no fraud)

Predicted fraud?

Was it actually a fraud?

	Y	**N**
Y	+/+ True positive	-/+ False negative
N	+/- False positive	-/- True negative

Confusion Matrix

of times that the corresponding possibility occurred

	Y	N
Y	2	4
N	8	6

Incorrect answers

Correct answers

How do We Evaluate the Quality of Our Classifier?

- **Precision:** How often a classifier is right when it says something is fraud?
- **Recall:** How much of the actual fraud we correctly detect

An aggressive detector will have high recall and low precision (catch most fraud, many false positives)

How do We Evaluate the Quality of Our Classifier?

- **Precision:** true positives / (true positives + false positives) = 2/10
  - True positives + false positives: this # tells us what % of the time an instance labeled positive was actually positive
- **Recall:** true positives / (true positives + false negatives) = 2/6
  - True positives + false negatives: total # of samples in which fraud actually occurred

Low recall (miss some fraud)
High precision (few false positives)

High recall (catch most fraud)
Low precision (many false positives)

Conservative
Flag fewer transactions

Aggressive
Flag more transactions

F-Score or F1-Score

- The harmonic mean of 2 numbers: combining 2 numbers that puts more emphasis on the smaller value

- F1 of x & y = (2 x precision x recall) / (precision + recall)

Example

```
from sklearn.metrics import confusion_matrix

from sklearn.metrics import classification_report

y_true = [0, 1, 2, 2, 2]

y_pred = [0, 0, 2, 2, 1]

target_names = ['class 0', 'class 1', 'class 2’]

confusion_matrix(y_true, y_pred)

print(classification_report(y_true, y_pred,
target_names=target_names))
```