Cost-based evaluation

Lecture 10

Outline

- Performance measure: error rate
- Generating test set
- Predicting performance interval
- Comparing two classifiers
- Cost-based evaluation

The Inadequacy of Accuracy

- As the class distribution becomes more skewed, evaluation based on accuracy breaks down.
  - Consider a dataset where the classes appear in a 999:1 ratio.
  - A simple rule, which classifies every instance as the majority class, gives a 99.9% accuracy – no further improvement is needed!

- Evaluation by classification accuracy also assumes equal error costs---that a false positive error is equivalent to a false negative error.
  - In the real world this is rarely the case, because classifications lead to actions which have consequences, sometimes grave.

Cost-based evaluation

- In practice, different types of classification errors often incur different costs

- The rare class is often denoted as positive
- The *confusion matrix*:

		Predicted class	
		Yes	No
Actual class	Yes	True positive	False negative
	No	False positive	True negative

Terminology

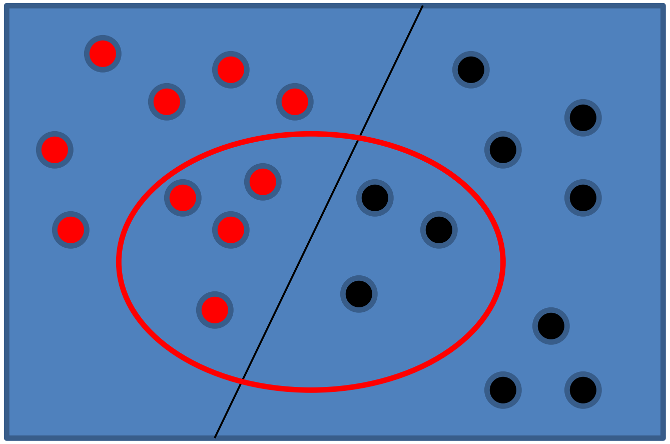
- The *confusion matrix*:

		Predicted class	
		Yes	No
Actual class	Yes	True positive	False negative
	No	False positive	True negative

True positives (TP) – the number of positive examples correctly predicted as positives
False negatives (FN) – the number of positive examples wrongly predicted as negatives
False positives (FP) – the number of negative examples wrongly predicted as positives
True negatives (TN) – the number of negative examples correctly predicted as negatives

Terminolgy. Fractions

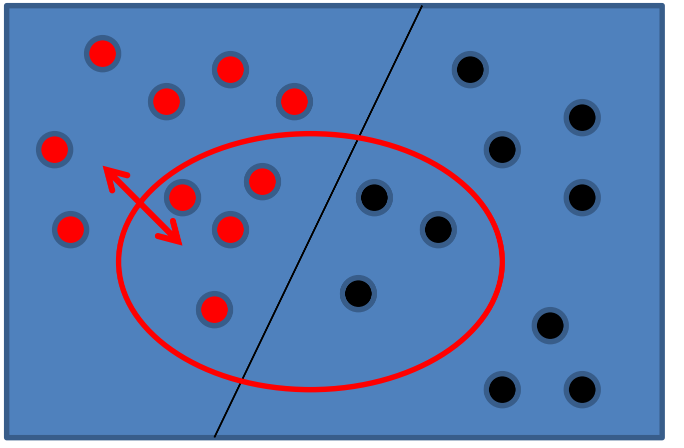
- Suppose you want to know what are all positive instances in your dataset (red dots)
- The classifier outputs as positives the instances inside the oval



Terminology. TPF, sensitivity or recall

- Suppose you want to find all positive instances in your dataset (red dots)
- The classifier outputs as positives the instances inside the oval

- True positive rate (fraction):
  TPF=TP/all positives
- In the example: 4 red dots out of 10 red dots – TPF=0.4
- Also called: sensitivity or recall
- High sensitivity or high recall mean that classifier found most of the relevant positive instances



Examples:
high-sensitive HIV test- if the person is sick, it will be diagnosed with high-probability

High-recall document query: the query brought most of the relevant documents

Terminology. Precision

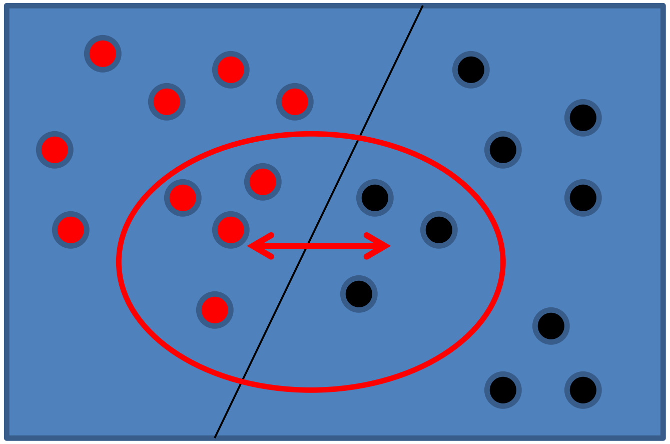
- Suppose you want to find all positive instances in your dataset (red dots)
- The classifier outputs as positives the instances inside the oval

- Precision (fraction):

  precision=TP/(TP+FP)
- In the example: 4 red dots out of 7 total dots which are all identified as positive

Precision=4/7

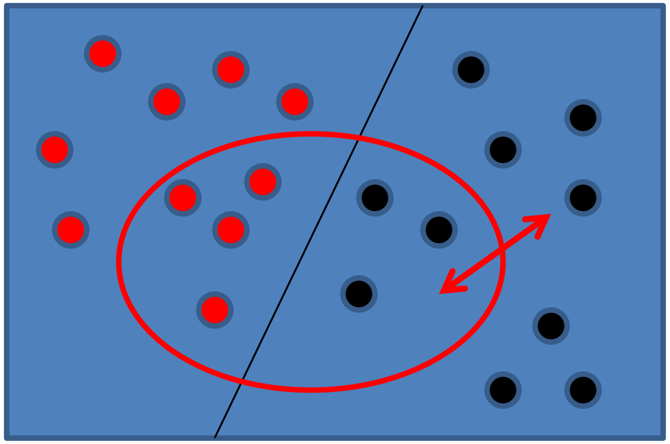
- High precision means that classifier returned more relevant results than irrelevant



Highly precise HIV test – whoever is classified as HIV-positive is most probably sick

Terminology. False Positive Fraction

- Suppose you want to find all positive instances in your dataset (red dots)
- The classifier outputs as positives the instances inside the oval



- False Positive Rate(fraction):
  FPF=FP/(all negatives)
- In the example: 3 black dots out of 10 total dots which represent all negative instances

FPF=3/10

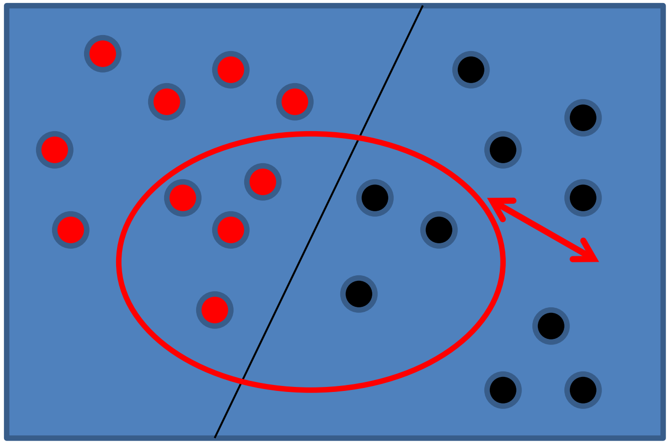
- High FPF means that classifier is not very specific – it brings a lot of irrelevant results

Example: mammography

If the person is diagnosed, it is not very likely to be really sick

Terminology. Specificity

- Suppose you want to find all positive instances in your dataset (red dots)
- The classifier outputs as positives the instances inside the oval



- Specificity (fraction):

  specificity=TN/(all negatives)
- In the example: 7 black dots which are left outside of the positive prediction out of total 10 negative instances

Specificity=7/10

- High specificity means that if classifier identifies something as positive, it is a high probability that it is indeed positive

Highly-specific test means that it is very low probability to be classified as positive, if the person is indeed negative

Specificity + FPF=1.00

Counting the cost. Example

		Predicted class	
		Class +	Class -
Actual class	Class +	-1	100
	Class -	1	0

Cost matrix

For example, HIV diagnostic test

- A cost matrix encodes the penalty of classifying records of one class as another. A negative value represents an award for making a correct classification

Counting the cost. Example

		Predicted class	
		Class +	Class -
Actual class	Class +	-1	100
	Class -	1	0

Cost matrix

		Predicted class	
		Class +	Class -
Actual class	Class +	150	40
	Class -	60	250

Confusion matrix for Classifier A

		Predicted class	
		Class +	Class -
Actual class	Class +	250	45
	Class -	5	200

Confusion matrix for Classifier B

The total cost of model A=150*(-1)+60*1+40*100=3910
The total cost of model B=250*(-1)+5*1+45*100=4255

Counting the cost. Example

		Predicted class	
		Class +	Class -
Actual class	Class +	-1	100
	Class -	1	0

Cost matrix

		Predicted class	
		Class +	Class -
Actual class	Class +	150	40
	Class -	60	250

Classifier A

		Predicted class	
		Class +	Class -
Actual class	Class +	250	45
	Class -	5	200

Classifier B

The total cost of model A=150*(-1)+60*1+40*100=3910

The total cost of model B=250*(-1)+5*1+45*100=4255

Cost matrix for different domains

- HIV diagnostic test

		Predicted class	
		Class +	Class -
Actual class	Class +	-100	10000
	Class -	10	0

Person dies untreated and infects others

Cost of additional testing plus some discomfort

Cost matrix for different domains

- Promotional mailing

		Predicted class	
		Class +	Class -
Actual class	Class +	-1000	1000
	Class -	1	0

Looses potential revenue

Cost of mailing

Cost matrix for different domains

- Loan decisions

		Predicted class	
		Class +	Class -
Actual class	Class +	-100	10
	Class -	50	0

Looses potential revenue

bancrupcy

Cost matrix for different domains

- Fault diagnosis

		Predicted class	
		Class +	Class -
Actual class	Class +	-10	1
	Class -	50	0

Additional test

System failure

Cost-based classification

- Let $\{\mathbf{p},\mathbf{n}\}$ be the positive and negative instance classes.
- Let $\{\mathbf{Y},\mathbf{N}\}$ be the classifications produced by a classifier.
- Let $c(\mathbf{Y},\mathbf{n})$ be the cost of a false positive error.
- Let $c(\mathbf{N},\mathbf{p})$ be the cost of a false negative error.

- For an instance E,
  - the classifier computes $p(\mathbf{p}|E)$ and $p(\mathbf{n}|E)=1- p(\mathbf{p}|E)$ and
  - the decision to emit a positive classification is

$$[1-p(\mathbf{p}|E)]*c(\mathbf{Y},\mathbf{n}) < p(\mathbf{p}|E) * c(\mathbf{N},\mathbf{p})$$