Fundamental Methods of Data Science

Class 21

Evaluating Classification Models

- Which metric have we been using for model evaluation?

Evaluating Classification Models

- Which metric have we been using for model evaluation?

- Accuracy (or error-rate) are common in evaluation
  - $\texttt{Accuracy} = \frac{\texttt{Number of correct decisions made}}{\texttt{Total number of decisions made}}$
- Can you think of a scenario where accuracy will not be as useful?
  - I.e. a model of lower accuracy will be preferred

Evaluating Classification Models

- Which metric have we been using for model evaluation?

- Accuracy (or error-rate) are common in evaluation
  - $\texttt{Accuracy} = \frac{\texttt{Number of correct decisions made}}{\texttt{Total number of decisions made}}$
- Can you think of a scenario where accuracy will not be as useful?
  - I.e. a model of lower accuracy will be preferred
- Medical examination: positive means a person has a certain disease
  - Model 1: 10 errors: 5 false positive and 5 false negative
  - Model 2: 15 errors: 15 false positive and 0 false negative
- Where
  - False positive means we wrongly predicted true for this instance
- which model is better?

Confusion Matrix

- A possible improvement to counting accuracy
  - Count different types of errors made by the classifier

Confusion Matrix

- A possible improvement to counting accuracy
  - Count different types of errors made by the classifier
- A confusion matrix counts correct and incorrect classifications
- A confusion matrix for a binary classification problem

	p	n
Y	True positives	False positives
N	False negatives	True negatives

- Positive/Negative refer to the actual class
- Yes/No refer to the answer our model gives
- What is the meaning of each diagonal?
- What is the matrix of the previous example?

Bad Positive and Harmless Negative

- In the previous medical example, we referred to being ill as positive
- Being positive normally refers to meriting special attention
  - Having a certain disease
- It also refers to a rarer event
  - Churning

Unbalanced Class Distribution

- Consider the following scenario for the churn problem
  - Analyst A gives you a model of 80% accuracy

	Positive	Negative
Yes	300	0
No	200	500

  - Analyst B gives you a model of 64% accuracy

	Positive	Negative
Yes	100	360
No	0	540

- What is the difference?

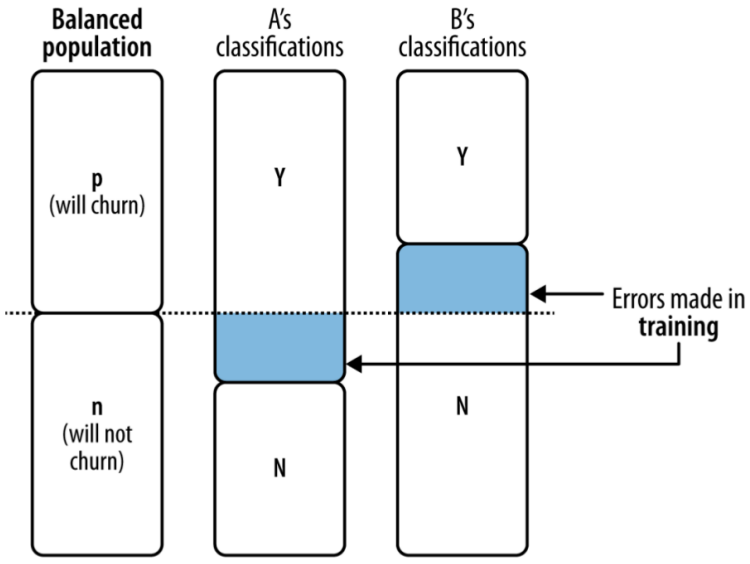
Unbalanced Class Distribution

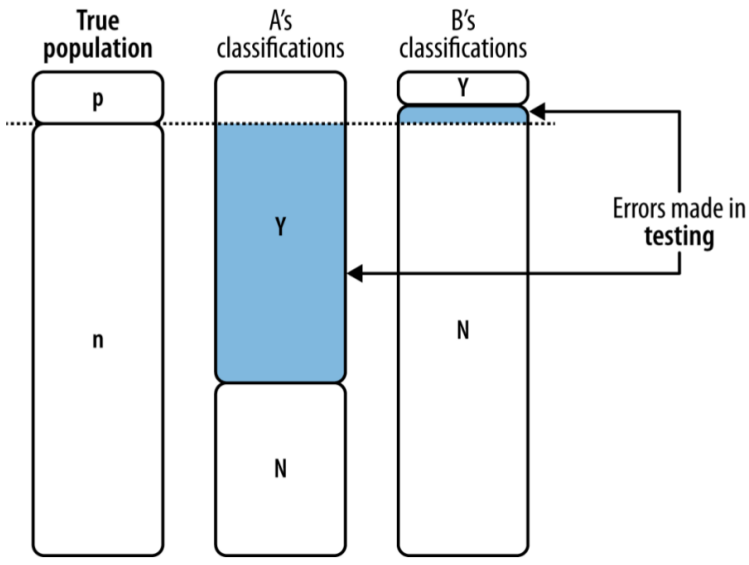
- The two analysts used data of different distributions
  - Analyst A used data with balanced distribution
  - Analyst B used data with representative distribution

	Positive	Negative
Yes	100	360
No	0	540

- What is the representative distribution?
- What would be analyst B accuracy on a balanced distribution?
- Can you see an easy way to get more than 80% accuracy on the representative distribution?

Unbalanced Class Distribution





Unequal Costs and Benefits

- How much do we care about the different **errors** and correct decisions?
  - Classification accuracy makes no distinction between **false positive** and **false negative** errors
  - In real-world applications, different kinds of errors lead to different consequences!
- Examples for medical diagnosis:
  - a patient has cancer (although he does not)
    - **false positive** error, expensive, but not life threatening
  - a patient has cancer, but she is told that she has not
    - **false negative** error, more serious
- Errors should be counted separately
  - Estimate cost or benefit of each decision

Evaluation Metrics

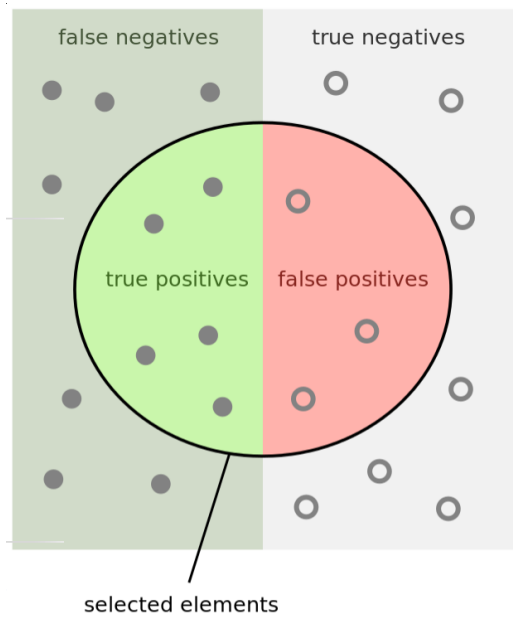
- Accuracy

```
score = sklearn.metrics.accuracy_score(y_true, y_pred)
```

- Confusion Matrix

```
matrix = sklearn.metrics.confusion_matrix(y_true, y_pred)
```

Precision and Recall



- $\texttt{Precision} = \frac{\texttt{TP}}{\texttt{TP + FP}}$
- $\texttt{Recall} = \frac{\texttt{TP}}{\texttt{TP + FN}}$
- $\texttt{F1-score} = 2 * \frac{\texttt{Precision*Recall}}{\texttt{Precision+Recall}}$

Evaluation Metrics

- Precision, Recall and F-score

```
>>> 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']
>>> print(classification_report(y_true, y_pred))
              precision    recall  f1-score   support

     class 0       0.50      1.00      0.67         1
     class 1       0.00      0.00      0.00         1
     class 2       1.00      0.67      0.80         3

avg / total       0.70      0.60      0.61         5
```

Workout

- For the Churn data can you get a confusion matrix for one of your models?
  - `http://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html`
- Is the data balanced or representational?

Analytical Framework: Expected Value

- The **expected value** computation provides a framework that is useful in organizing thinking about data-analytic problems
- It decomposes data-analytic thinking into:
  - The structure of the problem
  - The elements of the analysis that can be extracted from the data
  - The elements of the analysis that need to be acquired from other sources

Analytical Framework: Expected Value

- The **expected value** computation provides a framework that is useful in organizing thinking about data-analytic problems
- It decomposes data-analytic thinking into:
  - The structure of the problem
  - The elements of the analysis that can be extracted from the data
  - The elements of the analysis that need to be acquired from other sources
- The structure of the problem is given by the following formula

$$\texttt{EV} = p(o_1)v(o_1) + \ldots + p(o_n)v(o_n)$$

Expected Values of a Classification

- Consider the churn problem
  - What are the possible outcomes for a customer?

Expected Values of a Classification

- Consider the churn problem
  - What are the possible outcomes for a customer?
  - Can we get the probabilities of each outcome for a specific customer? How?

Expected Values of a Classification

- Consider the churn problem
  - What are the possible outcomes for a customer?
  - Can we get the probabilities of each outcome for a specific customer? How?
  - What is your estimation for the probability of churning on a representative distribution?

Expected Values of a Classification

- Consider the churn problem
  - What are the possible outcomes for a customer?
  - Can we get the probabilities of each outcome for a specific customer? How?
  - What is your estimation for the probability of churning on a representative distribution?
  - Should we target customers of 10% churning? 20%? How can well decide?

Expected Values of a Classification

- We can use the EV framework!
- What is missing for us to use the expected value formula?

$$\texttt{EV} = p(o_1)v(o_1) + \ldots + p(o_n)v(o_n)$$

Expected Values of a Classification

- We can use the EV framework!
- What is missing for us to use the expected value formula?

$$\texttt{EV} = p(o_1)v(o_1) + \ldots + p(o_n)v(o_n)$$

- How can we obtain these values?

Expected Values of a Classification

- We need some business understanding
  - What costs are involved with each outcome
- Responding to a product campaign example
  - Assume for a specific customer you got the following probabilities
    - $p(o_1) = 0.05$
    - $p(o_2) = 1 - p(o_1) = 0.95$
  - Assume having the following business understanding
    - Price of product: \$200
    - Costs of product: \$100
    - Contacting a customer: \$1
- Should we contact this customer?

Expected Values of a Classification

- We need some business understanding
  - What costs are involved with each outcome
- Responding to a product campaign example
  - Assume for a specific customer you got the following probabilities
    - $p(o_1) = 0.05$
    - $p(o_2) = 1 - p(o_1) = 0.95$
  - Assume having the following business understanding
    - Price of product: \$200
    - Costs of product: \$100
    - Contacting a customer: \$1
- Should we contact this customer?
- Compute what is the probability threshold

Expected Value of a Model

- We have seen how EV can be used with an existing model.
- We have also seen that our evaluation of models using accuracy score might sometimes not be very useful.
- We want to use the same framework, I.e. to use

$$\texttt{EV} = p(o_1)v(o_1) + \ldots + p(o_n)v(o_n)$$

- What are the possible outcomes now?
  - Hint: consider confusion matrices

Using the Confusion Matrix

- Data mining can give us a confusion matrix

	Positive	Negative
Yes	100	360
No	0	540

- We now have our outcomes, what are the probabilities of each?

Using the Confusion Matrix

- Data mining can give us a confusion matrix

	Positive	Negative
Yes	100	360
No	0	540

- We now have our outcomes, what are the probabilities of each?

- Assume having the same business understanding
  - Price of product: \$200
  - Costs of product: \$100
  - Contacting a customer: \$1
- What are the values of each outcome?

Using the Confusion Matrix

- Data mining can give us a confusion matrix

	Positive	Negative
Yes	100	360
No	0	540

- We now have our outcomes, what are the probabilities of each?

- Assume having the same business understanding
  - Price of product: \$200
  - Costs of product: \$100
  - Contacting a customer: \$1
- What are the values of each outcome?

- This matrix is called **Cost-benefit Matrix**