Lecture 12: Machine Learning and Naive Bayes

Key Questions

- What is machine learning?
- What is a "feature vector"? Where do they come from?
  - What's a feature vector we can get from a sentence of words?
- What is the difference between supervised and unsupervised learning?
- What is the difference between conditional and joint probability?
- What makes a probabilistic model probabilistic?
- What makes Naive Bayes "naive"?

Notes

Probability The probability of an event happening, written `p(event)`

Probability distribution Assigns probability values to all possible outcomes of one or more events, often written as a table; columns must add up to 1

Relationship between distributions $p(x, y) = p(x) * p(y|x)$

Bayes's Rule $p(label|data) = p(label) * p(data|label)/p(data)$

Naive Bayes Using Bayes's rule to predict $p(label|data)$ using observations of $p(data|label)$ from a training set.

Your Questions