## CS4025/CS5057: Continuous Assessment 1 (Classifying Emotion)

## Philip Hale

November 13, 2013

## 1 Data partitioning

Source: p. 187 of S&LP

- 10 test set (to evaluate the performance of the algorithm)
- 10 dev set (to fine-tune the algorithm during development)
- 80 training set (to 'teach' the algorithm).

Creating a dev set allows us to make adjustments to the algorithm while ensuring we do not bias towards the test data.

## 2 Testing

```
Comparison with existing implementation: http://classifier.rubyforge.org/

classifier = Classifier::Bayes.new "Angry", "Fearful"

# train
File.open('angry.txt').each { |grr| classifier.train_angry grr }
File.open('fearful.txt').each { |eek| classifier.train_fearful eek }

# test
classifier.classify "I hate you!"
=> Angry
classifier.classify "Please don't hurt me"
=> Fearful
```