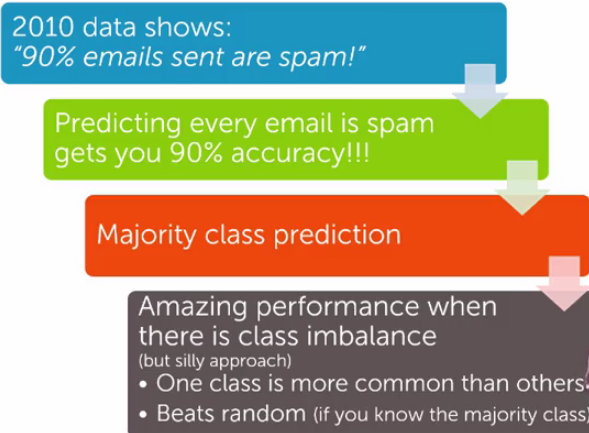
**### Evaluate Classification models**

* Error = # of mistakes / total # of sentences
  + Best possible value = 0.0
* Accuracy = # of correct / total # of sentences
  + Best possible values = 1.0

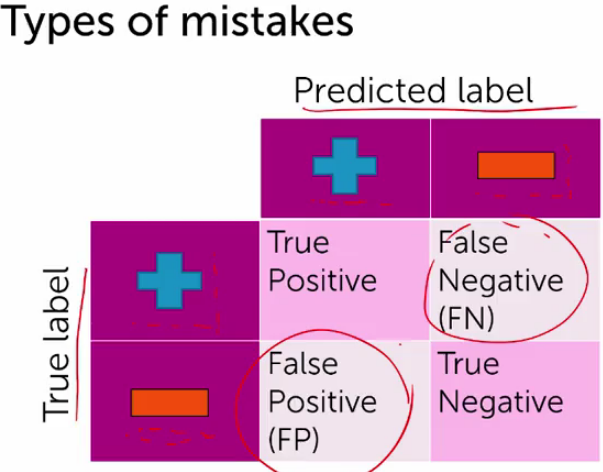
**### Measure accuracy**

* Binary classification: accuracy = 0.5 (a good model has accuracy > 0.5)
* For k classes: accuracy = 1/k
  + 0.333 for 3 classes, 0.25 for 4 classes
* Is classifier with 90% accuracy good?



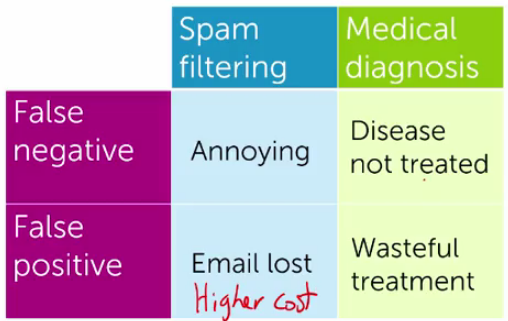
* Always digging in and ask hard questions:
  + Is there class imbalance?
  + How does it compare to a simple, baseline approach?
    - Random guessing
    - Majority class
  + What accuracy does my application/model need?
  + What is good enough for my user’s experience?
  + What is the impact of the mistakes?

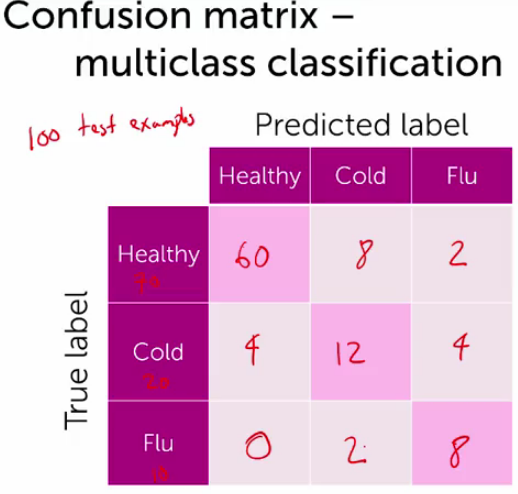
**### Confusion Matrix**



* Relationship between true label and predicted label
* True positive: true label is positive and predicted label is positive
* True negative: true label is negative and predicted label is negative
* False negative: true label is positive and predicted label is negative
* False positive: true label is negative and predicted label is positive

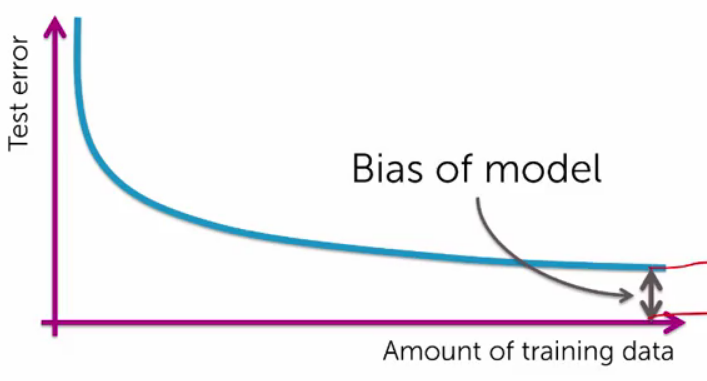
**### Cost of different types of mistakes**



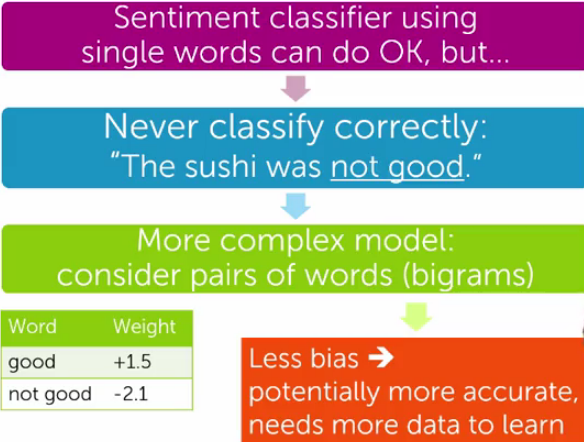


### Learning curves

* More data means lower test error.



* Limit: even with infinite data test error will not go to 0.
* More complex models tend to have less bias



### Predicting probabilities

* How confident are you in the prediction