* Induction
  + Generalize from a specific example to tell the future if you will
  + \*\*Carful because the past doesn’t always predict the future
  + \*\*Bias – any bases to choose among generalizations other than the data itself. Help us choose beyond just the data, comes from outside the data.
    - It is essential to learn
    - In the absence of bias you cannot learn.
* Deduction
  + Given a general principle tell the truth
* Unsupervised Learning
  + No external forces to help learning
* Supervised
  + Some external direction or help
* Bias – Variance Trade Off
  + Strong – number of generalizations decrease, decrease variance
  + Weak – number of generalizations increase, increase variance
* \* As humans we adapt and change our bias constantly. It is what allows us to learn.
* How do you know what you’re doing is any good…when you make a guess to the data?
  + Bias
    - Any basis for choosing one decision over another, other than strict consistency with past observations
* If you have no bias you cannot go beyond mere memorization
  + The power of a generalization system follows directly from its biases
  + w/o bias you can only memorize
* Unbiased Language -- has no bias – it can represent everything… any set can be given a name.
* Unbiased Generalization Procedure – it can pick any set…
* Version Space Algorithm
  + G – everyone
  + S – everyone we know to be evil
  + Everything else
* If you have an Unbias language and Unbias procedure – you cannot learn, you can only memorize.
* Sources of Bias in Learning
  + The representaitoin language cannot express all possible classes of observation
  + The generalization procedure
* Overfit – become so good at classifying test data you loose ability to correctly classify new data
  + Avoid
    - Stop early – either number or proportion reaches a limit
  + Prune it – POST Pruning
    - Hold some data back – then run it back through and see how the held data works
    - Cut a branch and see if it gets better… If it does cut the branch
* Questions:
  + 25-40
  + 50-300
  + 2-5
  + 5-30
  + 500-900
  + 1000-2000
  + 1600-1700
  + 60-120
  + 10000-20000
  + 150000-900000
* Why didn’t you guess –inf to inf?
  + Because we’re bias
  + And we hold onto our biases
    - Don’t want to look stupid
    - I’m close
    - I know the answers
* Information gain == difference in entropy
* == entropy before – entorypy after
* Why Entropy over Accuracy?
  + Entropy we just need to change the distribution
  + However if using accuracy…we are trying to change the majority with the split
    - This could be very hard to do

Eager Learner – once model is created you can discard the data.

Lazy Learner – can’t discard the data

Batch Learner / Inline or Incremental Learner

* \* know the biases of each algorithm
* K-Nearest Neighbor:
  + Bias: things that look like me should be classified like me.
  + Assumption : we can put things into some meaningful space and the distance around them is meaningful.
  + Assumption : we have continuous space
  + ….What is the distance between red and green?
    - How do you measure these things.
    - Could do ordinal – discrete but the difference makes sense.
* DTree – bias : build small trees, no data, set up model, re run to change
* K – bias: , incremental, lazy, needs data
* NB – Inbetween K, and Dtree
  + Uses model
  + But can update model easily
  + Doen’t have to keep the data
  + P(h|D) – find H that maximizes this thing
* Bayes
  + Way different, from the others. Its run on probabilities
  + Prier is a big deal…tells you how it changes as data becomes available. Probabilities drive the decisions you take.
    - List assumptions:
      * Congunction of att values
      * Att are independent given the class
        + NB is still robust
      * Est probabilities with frequencys
* -Reminders
  + Tension between the prior and the data
  + Handles new data easily
  + Incremental
* M-estimates
  + Get ride of the zero probalbilty thing… Like you will never play tennis when its rainy
* K- nearst
  + Bias – what looks like me should be classified like me
    - The distance metric
  + Incremental
  + Harder to explain why now vs Decision Tree
* Black Box stuff
* Sub-simbolic
* Perceptron
  + Must be linearly separable
    - Otherwise they cannot converge