Learning from Data lecture notes

Overfitting

* Introduction
  + Everyone can fit, professionals can prevent overfitting
* What is overfitting? Vs bad generalization
  + Overfitting happens by using functions that have too high
  + Bad generalization can happens when you can’t approximate target well
  + Similar thing can happen when just training models
    - Effective VC dimension increases as time goes on in neural networks
    - Hence difference increases between and (albeit both decreases mostly)
    - When starts going up as goes down, you’re overfitting
    - Detecting the above is probably important, we want to return the weights at minimum
  + Overfitting: fitting data more than warranted
    - Why is this bad? You’re fitting noise
    - That’s harmful – generalization is harmed
  + Overfitting to noise and to insufficient data generation
    - Both harmful to generalization
    - Latter occurs when you try fitting to 50th order polynomial with 15 data points (won’t accurately portray even though it’s noiseless)
    - Remember: pick model based on data resources, not model complexity
    - Pay attention to other facets (symmetry, etc) but not model type
  + Learning curves incorporate this:
    - includes both inability to approximate of target and noise
    - Even if a model can perfectly approximate target, for large ranges of you may still want to go with simpler model
* Noise experiment
  + Recommended for us to implement too
  + Use legendre polynomials (more interesting than polynomials where coefficients are randomly generated)
  + Relative overfit measure,
  + Outcomes:
    - Conventional noise isn’t all that can be overfit to
    - Logo for book comes from these output graphs
  + Noise due to randomness is stochastic noise
  + More complex target function-induced noise is called deterministic noise
  + What effects the above two?
    - More data, less noise for both
    - Increasing stochastic noise + deterministic noise, more overfitting
* Deterministic noise
  + Part of the target cannot capture,
  + Why is this noise?
    - Because we can’t capture it fully, the data misleads the learner if we attempt to capture it
  + Differences with stochastic noise
    - Depends on
    - fixed for a given
  + behave the same with stochastic noise in terms of machine learning algorithms
  + impact on overfitting
    - negatively impacts once target function is greater than learning model
    - learning models fit stochastic/deterministic noise with finite samples
* bias-variance
  + decomposition now has bias + var + expectation of
  + bias = deterministic noise
  + expectation of is stochastic noise
* dealing with overfitting
  + regularization (brakes)
  + validation (check bottom line)
* regularization intro
  + ­apply a slight brake (to prevent getting points right exactly), dramatically increases fit