Learning from Data lecture notes

Validation

* Validation vs regularization
  + Recall,
  + Regularization attempts to estimate with the regularizer
  + Validation attempts to estimate directly
  + estimates with variance (albeit a error mean of 0)
  + is the error with respect to a test set
    - Again, error is unbiased and variance is
    - O(x) = order of x
  + Why not increase of test set? You’re taking values from training
  + Impact:
    - points in validation, points overall
    - Small small estimate
    - Large increasing (recall learning curves)
  + Notation: is best model trained on training set
  + Tactic:
    - Pick best model with training set and test sets
    - Train best model on whole set and return that
  + Problem:
    - Estimate could be bad
    - Best hypothesis trained an all the data
    - Now: small k = estimated error will be small
    - Now: large k = large difference between error of and
  + Solution:
    - (rule of thumb)
  + Why is this validation?
    - The second you make decisions on test sets, you bias their results
    - They become validation sets to show that they are positively biased
    - Bias will be weak however assuming you don’t intensively use the validation sets (so it remains a good estimate for )
* Model Selection
  + Assume you have models, to be evaluated using to get with validation error
  + Experiment showing bias impact as increases
  + We should use for model selection some but not excess. How much is right?
    - Use it only to “train” on finalists models,
    - If is infinite, 2 parameters (when picking or early stopping number of epochs ) are okay both correspond to a VC dimension it is okay
    - If is infinite, 20+parameters are far in excess
  + Data contamination
    - Training sets = totally contaminated
    - Test sets = clean
    - Validation sets = slightly contaminated
      * Potentially have multiple validation sets
* Cross-validation
  + Recall , so we want to be small (for first approx. to hold true), and to be large (for second approx to be true)
  + Both can be achieved! LOOCV
    - Use data for training, for validation.
    - Gather with being each specific data point (so you train and times) (each model was trained with points)
    - , you now have each constructing a validation set of size
      * albeit not an independent set since point was trained on to predict , but impact is very minor (validation set of taking into account loss of power from no independence)
    - Don’t use validation CV for heuristic-based model selection, just pick on impact with respect to
  + -fold CV
    - Leave extra out with each iteration
    - Less training required
    - works nicely