CSE 446 Notes

Residual Sum of Squares (RSS): , (Least squares optimal solution)

Handling intercept: (1) Show that (2) “Demean” linear regression model so \* holds

Interpreting the result: When other features are fixed, the estimated change in for one-unit change of

Reason for minimizing RSS: if and which means , the MLE for produces minimum RSS.

Feature map: in polynomial regression, transform each data point to higher-dimensional feature vector

: jth feature associated with input (jth basis function)

Supervised Learning: learning a function that maps inputs to outputs based on labelled examples

For linear regression, the loss function is RSS.

Handling an intercept in linear regression: model is , , w still the same.

In the case of feature map, replace with

Training Error: , dependent on how loss function is defined. Training error is overly optimistic on training data (unless training set includes everything you might ever see).

Generalization error:

As model complexity goes up, training error goes down; generalization error goes down, then goes up.

Testing Error: same as training error, but on a labelled testing data set.

Three sources of error:

Data inherited noise: , irreducible.

Bias: Let , then -> is our model flexible enough to catch true ?

Variance: How much do specific fit varies from expected fit?

Low complexity: high bias, low variance High complexity: high variance, low bias

If and ,

Regularization: imposes “simpler” solutions by a complexity penalty

Rigid Regression: penalized least-squares regression. Trade of model complexity with training error

, solves to

Shrinkage Property: assume (a special case),

Assume and , , we have

, .

Leave-One-Out (LOO) cross validation: is the set of training data with data point j removed

is usually pessimistic because learning with less data points gives worse answer.

LOO is almost unbiased, so it can be used for model selection (i.e. picking the right in regularization).

k-fold cross validation: Divide data to k equal parts

Estimation error on set :

k-fold error: average over all sets

Faster to compute and much more pessimistic than LOO (because model is trained with even less data points)

Usually choose

Given a set of input data: (1) split to train and test set (2) use k-fold cross validation on training set to train predictor and choose magic parameters such as (3) Assessment: use test set to check the accuracy of the model

Theorem for Penalized Least Squares: For any for which achieves the minimum, there exists a such that subject to

LASSO Regression: , keep and for not dealing with offset terms.

Optimizing LASSO objective one coordinate at a time: For , and

Subgradient: , is a minimum if is a subgradient at

Coordinate Descent for LASSO (Shooting Algorithm): Pick a random coordinate , compute ,

L1 Regularization is one way to do variable selection (finding a sparse solution). LASSO is non-differentiable, but convex

Binary Classification:

Sigmond for Binary Class:

Linear Decision Rule:

Conditional Likelihood: if , , , is a convex function, no closed-form solution

To prevent overfitting: add a regularization term (i.e. )

Gradient Descent:

With Taylor Series Approximation:

For , ,

Total Derivative:

Stochastic Gradient Descent: Use one data point to estimate the actual averaged gradient over all coordinates

Each data point contributes to of regularization.

Estimated Derivative: for some random data point

SGD has noisy convergence.

Perceptron Algorithm: At each step, make prediction ; update

Max number of mistakes perceptron algorithm can make: in linear separable case, (: upper bound of norm of all ’s, is margin.)

Hinge Loss:

K nearest neighbor: as k increases, larger bias and smaller variance; as n approaches infinity, 1-NN rule error is at most twice than bayes error.

Hyperparameters tuning: grid search/random search/bayes optimization

PCA: Compute orthogonal vectors to maximize prpjected variance.

Kernel Trick: ,

Decision Tree Stopping: all data agrees on y, already split on all features, no split reduces classification error.