Linear Regression

* Supervised learning that assumes a linear relationship between predictor Y and features XP
* Useful for quantitative response

Simple Linear Regression

* Y = f(x) + e
* = B0 + B1x + e
* B represent intercept and slope
* E is the irreducible error
* A change in X by one unit is associated with an average change in Y by B-hat1 units

RSS:

* Sum of squared residuals
* Least Squares regression chooses coefficient by minimizing RSS

TSS:

* Total sum of squares
* Sum of each response subtracted by the mean response

Assessing accuracy of coefficients

* Uncertainty captured via standard error (SE)
* RSE = sqrt(RSS / (n-2))
* SE of coefficients are proportioanal to SE of regression and inversely proportional to square root of sample size
* Using standard errors of predictors, we can estimate 95% confidence interval
  + If we would calculate coefficients from 100 training sets, our estimates would fall within the range of the above intervals in 95 out of 100 samples
  + Slope: [B1 – 2SE(B1), B1 + 2SE(B1)]
  + Intercept: [B0 – 2SE(B0), B0 + 2SE(B0)]
* Can be used to perform hypothesis testing
  + H0: There is no relationship between X and Y
  + HA: There is some relationship between X and Y
* P-value:
  + T = (B1 – 0) / (SE(B1)
  + Probability of any value equal t or larger
* R2 = (TSS – RSS) / TSS = 1 – (RSS / TSS)

Multiple Linear Regression:

* Y = B0 + B1x1 + B2x2 + … BpXp
* Ideal scenario when predictors are uncorreleated
  + Balanced design
  + Each coefficient can be estimated and tested separately
  + Interpretations such as “a unit change in Xj is assosciated with a Bj change in Y, while all other predictors are held constant” is possible
* Correlations amongst predictors cause problems:
  + Variance of all coefficients tends to increase
  + Interpretations become hazardous
* Claims of casuality should be avoided
* Selecting predictors
* Best subsets regression
  + Compute least squares fit for all possible subsets and choose based on criterion that balances training error with model size
  + Forward selection:
    - Begin with null model – intercept but no predictors
    - Fit p simple linear regressions and add to null model the predictor that results in lowest RSS
    - Add to that model the predictor that results in the lowest RSS among all two-predictor models
    - Continuine until stopping rule is satisfied
  + Backward selection:
    - Start with all predictors in the model
    - Remove predictor with the largest p-value, that is, predictor that is the least statistically significant
    - New (p-1) predictor model is fit, predictor with largest p-value is removed
    - Continue until stopping rule is satisfied
    - Has place but has issues