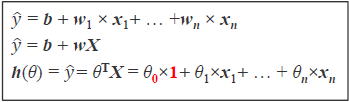
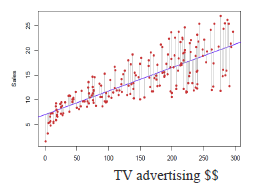
Linear Regression (LR) Concepts

* LR models must have numerical responses (i.e. dependent vars, target features)
  + LR estimates linear parameters (i.e line) to fit data to min residue or error
* Y = target (responses, dpdnt), X = features(predictors)
* Univariate LR = One X feature, Multivariate = 1+ X
* Build a regression model by computing coefficients Θ as

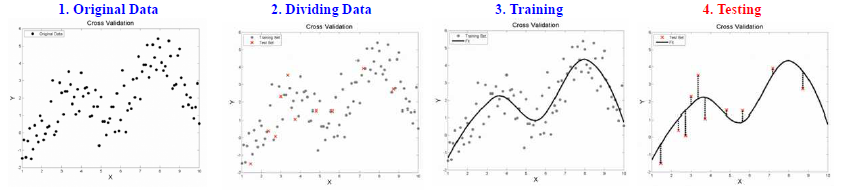
ŷ = ΘTX. Verify ŷ against Y.

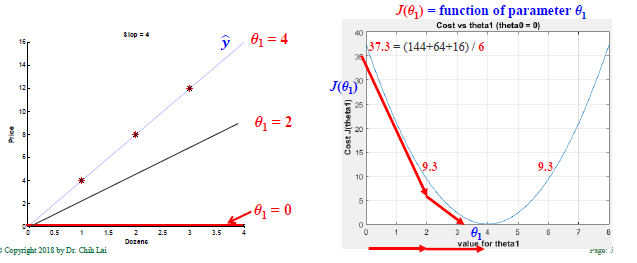
* Multivariate LR using weights for each predictor
* Categorical = “Classification”
* R-Square, R2 (Coefficient of Determination)
  + Smaller residual (LR fits better to data) comparing to avg🡪 R2 closer to 1. ()

Norm Plot can be used to spot potential outliers.

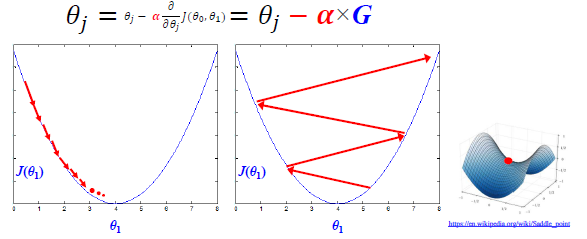
* Categorical variables (use dummy variables)
  + Continuous: age, income, salary
  + Categorical: gender, rank, department, city, etc.
* Regularization
  + Cross validation
  + Prepare original data
  + Divide data into training and test sets
  + Build a model using training data
  + Validate the model using test set



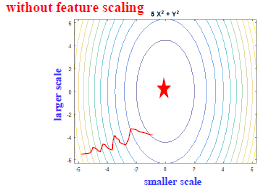
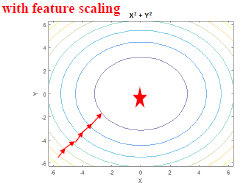
* + Gradient = minimize MSE
  + Visualizing Cost Function



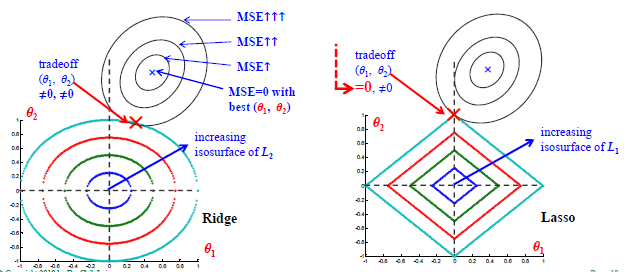
* Gradient descent (GD) may find local minimum, not global minimum.
* α is the learning rate = how big each downhill step.
  + α >0 must always hold
  + if α too small, GD takes long time to find best Θ
  + if α too big, GD overshoot local minimum and may fail.
    - Unstable learning

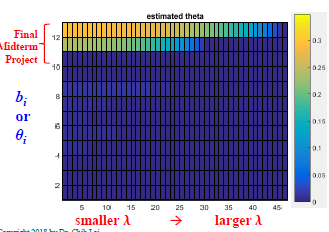
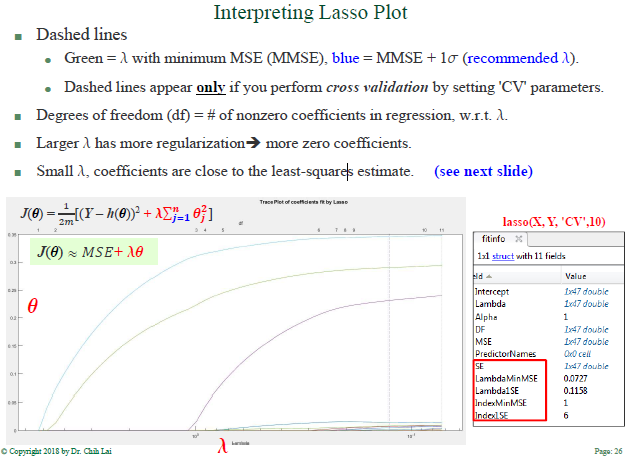


* Speedup Gradient Descent = Feature Scaling

* Use fminunc() to automate the GD process
* Normal Equation (XTX)-1XTy
  + No need to do feature scaling
  + No need to choose α, no need for iteration, slow for large dataset
  + Works for regression, but NOT work for classification
* Regularization
  + Training Accuracy ^ Validation accuracy ^
  + Training Accuracy ^ Validation Accuracy ˅ (overfitting stp training)
  + Training Accuracy ˅ Validation Accuracy ^
  + If test accuracy is not good, how to improve it
  + Remove outliers, avoid over fitting
  + Traditional Feature Selection
    - Entropy, p-Value, Sequential Covering methd
    - Test ALL variables on at a time and gradually add/remove to/from final model
    - Variables are either included or excluded, no partial inclusion
    - Retain only important predictors to avoid overfitting
    - Regularization algorithms include
  + Ridge regression
    - Shrink Θ but Not completely 0, compensate for correlated features w/o eliminating them
    - Not as effective as Lasso in eliminating coefficients.
  + lasso regression (Least Absolute Shrinkage and Selection Operator)
    - Drives some Θ toward 0 relatively quick, great for feature selection w/ wide data sets.
    - Wide data sets = ^ attributes ˅ records

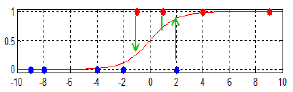


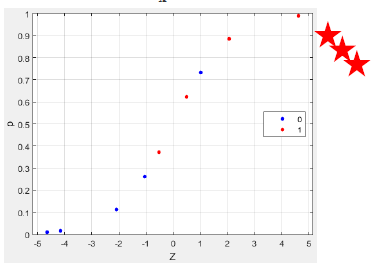
* + Lasso Identifies unnecssry predictors
  + 
  + 

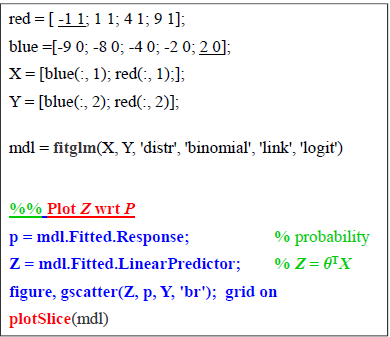
***Logistic***

Logistic regression is a model where the dependent var is categorical.









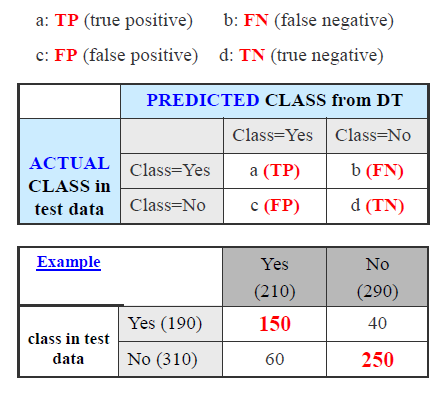
Θs tell us which attributes have positive or negative impact on P or classification.

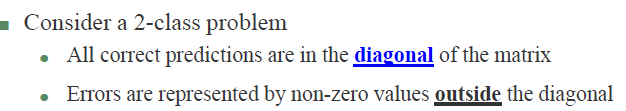
Regularization

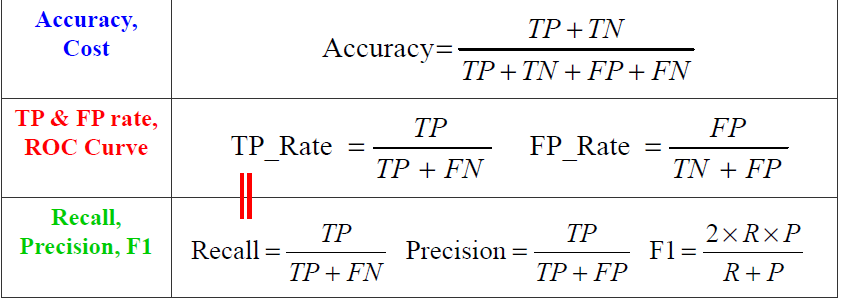
* Reduce the # of predictors and identify important ones
* Shrink Θs, potentially avoid ***overfitting*** problem

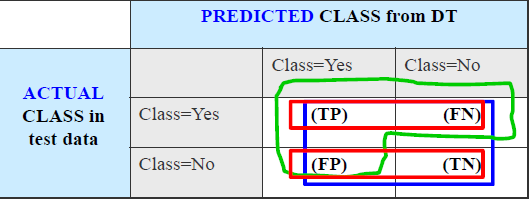
***Quality***

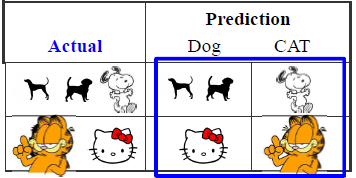
Confusion Matrix

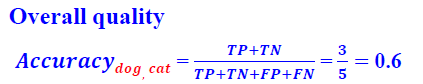


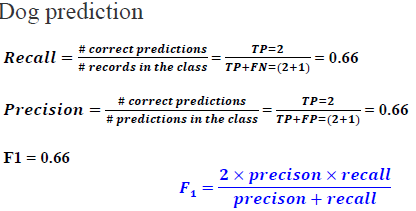


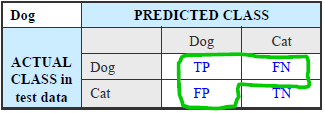


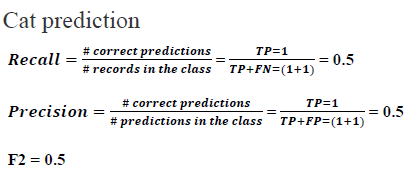


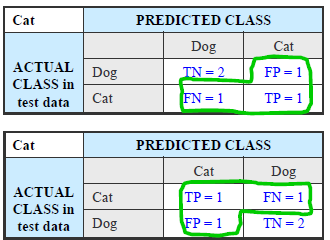




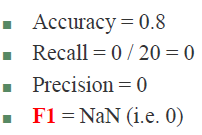


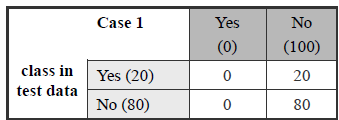


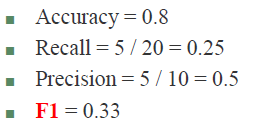


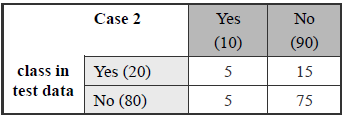


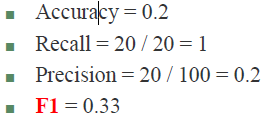


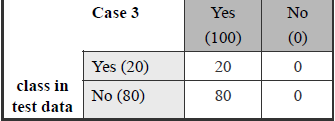




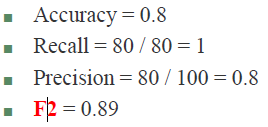


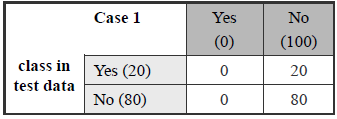


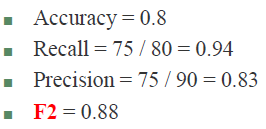


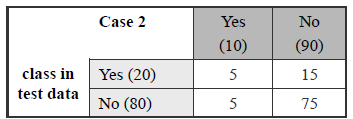


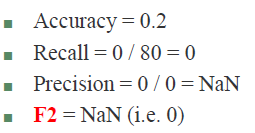


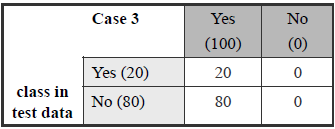












2 scientist walk into an elevator.

Balance = use accuracy

Not balance = don’t use accuracy