* Perceptron
* Linear regression
* Logistic regression
* Grading descent
* Decision tree
* kNN
* cross validation
* multi class classification
* Support vector machine
* Soft SVM
* Kernels
* **Probabilistic modeling**
  + Choose restaurants based on features
  + Gaussian distribution, when joint/conditional/marginal probability is gaussian
  + do MLE on gaussian with respect to means, variance
* Naïve Bayes theorem
  + Joint probability of data
* Gaussian discriminative analysis
  + Models the class conditional distribution directly as multivariate Gaussian
  + P(x,C)=NB
* Gaussian generative modeling
  + model class-conditional distributions P(x|C) and then class priors P(C) to get to class-conditional
* Maximum likelihood estimate
  + Maximize the joint probability by taking the negative log of an objective function
  + Take derivative of the log expression and set to zero to obtain an expression for the parameter that we are maximizing
* **Unsupervised learning**
* K-means clustering
  + partition data ser into K clusters with u\_k representing the kth cluster.
  + minimize objective function representing r\_nk\*distance(a point, mean) where r\_nk is 1 when the point is in k cluster.
  + add regularization term: take derivative(gradient) to minimize J
  + change L2 distance, take derivative to minimize J
  + subgradient/subdifferential (of absolute value)
    - subdifferentiable
    - to show x\* is a minimizer of a convex function:
      * the function is subdifferentiable at x\*
      * the subgradient at x\* includes 0
  + Image compression
  + Entropy function
* Principal component analysis
  + Eigenvalue/eigenvector
  + Data covariance matrix
  + Eigenvector corresponding the largest eigenvalue: maximize the variance of the projection onto a 1D dimensional space
  + reducing dimension from to D to M
* Expectation maximization using Gaussian mixture model
  + Soft K-means, P(x\_n) = probability of the nth point in kth clusters
* Ensemble methods
  + Boosting
  + Bagging
* Neural networks
* Number of parameters
  + logistic regression 2M+1
  + GDA 1+2M+M(M+1)/2
  + GDA with NB assumption 1+3M