# The weeks

+ the score

+ the links to project 1 and project 2

+ this review of 5 weeks

+ the answers for test 4

### Week 1a: linear/logistic models, training/testing, Gradient Descending

+ linear model is to capture the linear relationship between two variable x and y, or we can write y = wx + b for prediction

+ logistic model is to capture the probability of Bernoulli distribution of y and w\*x + b for binary classification.

+ learning the model by gradient descending.

+ training set is to build the model and test if the model fit; testing set is to test if the model can handle new input.

### Week 1b: mixture model, GMM, kmeans, EM

+ mixture model is to combine simple distribution to capture complicated dataset.

+ Gaussian mixture model is a mixture model with Gaussian distributions as mixture components.

+ using kmeans or EM to learn GMM

### Week 2a: SVM

+ to classify by separating the samples with line w\*x + b

+ w can be built based on the support vectors.

+ we learn support vectors by SMO

+ we can use kernel to classify non-linear datasets (by transforming feature with kernel).

### Week2b: boosting, bagging, random forest

+ combining methods

+ we have to tell the differences between the methods

+ bagging: to remove (bootstrap) the rows only

+ random forest: to remove (bootstrap) the rows and columns and growing tree

+ boosting: to build the new classifier based on the error sample only (emphasizing on error)

+ tree: to divide the samples into homogeneous regions to reduce error (more homogeneous, fewer errors). Tree is very fast classifier.

### Week 3b: HMM

+ to capture the relationship between symbol (observed) sequences and the state sequences.

+ for state Y and symbol sequence X, we will have the model p(X, Y) given by (A, B, π).

+ A is the probability state transition matrix. A(i,j) is the probability that we move from state i to state j.

+ B is the emission matrix. B(j, k) is the probability that state j produces symbol k. When we hear the cellphone number in the first place j=1 we expect to hear the word k=0 and the second j=2 we expect to hear k=9.

+ π(j) is the probability that state j will be the initial state (or the starting point of evaluation).

+ learning the HMM model with Baum-Welch procedure (E-M procedure with different E step and M step); Viterbi training algorithm

+ find the best state sequence given the symbol sequence: Viterbi

### Week 3b: sampling, optimization

+ sampling method: to draw samples from a distribution, if we draw enough, we can reconstruct the original distribution.

+ some methods we have learn: uniform U(a, b), box-muller for N(mu, sigma), 12U-6 for N(mu, sigma), rejection for complicated distribution, importance sampling also for complicated distribution without rejection.

+ exploring methods (to find a path from initial point to the optimal point): gibbs sampling, simulated annealing, enumeration, uniform sampling.

+ optimization: to find the best (or the optimal, or the maximal, or the minimal) point of the function f(x) given the range or the constraint x ∈R

+ kmeans, gradient descending, gibbs sampling, simulated annealing are all to find the optimal solution from one initial point.

### Week 4: clustering

+ different methods to divide the samples into groups (clusters): kmeans, GMM, hierarchal clustering, spectral clustering.

+ hierarchical clustering: look at the tree and read it.

+ to evaluate the error for kmeans.

+ to understand the E-M process (initializing, labeling, averaging)

+ to tell the differences between EM and kmeans

### Week 5: deep learning

+ motivation to deep learning?

+ concept of latent/hidden variables and how to work with them

+ from model to equations and from equations to model

+ the 3 typical types of DL model