Topic

- ☐ The midterm covers all topics so far discussed in class, in the assigned readings and in the homework. The goal of this outline is to focus your studying on the major topics that will appear on the midterm. You should be familiar with the mathematical concepts covered in the course, including the structure of the major objective functions and optimization methods. However, we do not expect you to memorize formula or complete complex derivations.
- ☐
- ☐ Machine Learning Foundations
  - ☐ Supervised, unsupervised, semi-supervised, reinforcement learning
  - ☐ classification vs. regression
  - ☐ linear data, linearly separable
  - ☐ generalized linear models
  - ☐ hypothesis classes
  - ☐ loss functions
  - ☐ cross validation, train/dev/held out data
  - ☐ hyper-parameter tuning
  - ☐ minimize/maximize functions, optimization, learning rate/step size
  - ☐ linear prediction functions
  - ☐ Bias variance tradeoff
    - ☐ Regularization, l1, l2
    - ☐ over-fitting, under-fitting
- ☐ Linear Regression
  - ☐ fitting regression functions
  - ☐ least squares linear regression
- ☐ Logistic Regression
  - ☐ data likelihood
  - ☐ relation to linear regression
  - ☐ Regression for classification
- ☐ Perceptron
  - ☐ geometric interpretation data
  - ☐ convergence
  - ☐ stochastic vs. batch optimization
  - ☐ online learning framework, order of the stream
  - ☐ assumptions behind online methods vs. batch methods
  - ☐ model combinations
  - ☐ neural view
- ☐ SVM
  - ☐ max margin principle, functional margin, geometric margin
  - ☐ Form of the optimization problem
  - ☐ Comparison to logistic regression, perceptron
  - ☐ online SVM (Pegasus)
  - ☐ dual/primal relations
  - ☐ slack variables and non-separable data
- ☐ Kernels
  - ☐ linear vs. non-linear classifiers
  - ☐ kernel trick
  - ☐ examples of kernels
  - ☐ issue of over-fitting
  - ☐ definitions of kernels
  - ☐ learning non-linear functions
  - ☐ kernel versions of other classifiers
- ☐ Decision Tree
  - ☐ learning algorithm, base cases
  - ☐ Good trees vs. bad trees
  - ☐ advantages and limitations
  - ☐ information theory
- ☐ Boosting
  - ☐ Ada-boost
  - ☐ weak/strong hypotheses
  - ☐ explanations for why boosting is good
  - ☐ Practical issues of boosting
- ☐ Neural network
  - ☐ network structure
  - ☐ non-linearities
  - ☐ activation functions
  - ☐ hypothesis classes
  - ☐ convex and non-convex
  - ☐ classification and regression objectives
  - ☐ forward and back propagation, chain rule, gradients
  - ☐ multi-layer perceptrons
  - ☐ different network structures
- ☐ Deep learning
  - ☐ motivation
  - ☐ Deep network training issues
  - ☐ learning representations
  - ☐ auto-encoders
- ☐ K-nearest neighbors
  - ☐ general algorithm

Topic

- ☐ behavior as K is varied
- ☐ multi-class uses of KNN