

Similarities/Differences Between algorithms:

- 1 Nearest Neighbor (1-NN)
 - k Nearest Neighbors (k-NN)
 - Decision Trees (Decision Stumps, etc.)
 - The Perceptron
 - Logistic Regression
 - Polynomial Regression (and other feature expansions)
 - Support Vector Machines (SVMs), Kernels
 - Ensemble Methods, AdaBoost
-
- How will these algorithms perform on various datasets?
 - What will decision boundaries look like?
 - Global vs local optima during learning?

Theory:

- Training vs. Testing, Generalization
- Overfitting, Underfitting
- Bias vs. Variance
- Cross Validation
- Shattering, VC Dimension
- Accuracy, Precision, Recall

Decision Trees (working knowledge):

- Splitting
- Entropy / Conditional Entropy
- Information Gain
- Pruning (deep vs shallow / stumps)
- Overfitting

Perceptron, Linear Regression, Logistic Regression:

- Regularization (relation to bias and variance)
- Learning, gradient descent
- Loss functions
- p-Norms

Support Vector Machines

- What is C ? (relation to bias and variance)
- What are support vectors, margin (how does changing support vectors change decision boundary/margin)
- Kernels / inner products (Linear, Polynomial, Gaussian)

Ensembles

- AdaBoost (bias / variance, when do you stop training)
- Properties of individual learners (what can they be, high/low accuracy)