Similarities/Differences Between algorithms:

- Decision Trees (Decision Stumps, etc.)
- Linear Regression (gradient descent)
- Polynomial Regression (and other feature expansions)
- The Perceptron
- Logistic Regression
- Neural Networks
- 1 Nearest Neighbor (1-NN)
- k Nearest Neighbors (k-NN)
- Ensemble Methods, AdaBoost
- Support Vector Machines (SVMs)
- Naive Bayes

- How will these algorithms perform on various datasets?
- What will decision boundaries look like?
- Global vs local optima during learning?
Supervised Learning:

- Difference with Unsupervised Learning, & Reinforcement Learning
- Machine Learning algorithms \(<P,T,E>\)
- Classification Vs. Regression
- Error, Accuracy, Confusion Matrix, Precision, Recall

Theory:

- Training vs. Testing, Generalization
- Overfitting, Underfitting
- Bias vs. Variance
- Cross Validation
- Shattering, VC Dimension
Perceptron, Linear Regression, Logistic Regression:

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

Neural Networks

- Structure (layers)
- Activation functions (linear, logistic)
- Backpropagation

Decision Trees (working knowledge):

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

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

Support Vector Machines

- What are support vectors?
- margin (how does changing support vectors change decision boundary/margin)

Naive Bayes

- Basic probability, Bayes Rule
- Prior, likelihood, posterior
- Independence assumption and why does it help?
Randomized Optimization:

- Local Search / Gradient Descent
  - Drawbacks
  - Random Restarts
- Simulated Annealing
- Beam Search
- Genetic Algorithms

Unsupervised Learning:

- Difference with Supervised Learning, & Reinforcement Learning
- Clustering vs. Feature Selection vs Dimensionality Reduction
Clustering:

- K-means Clustering
  - Importance of similarity measure
  - Should have a working knowledge of this algorithm
- Agglomerative / Hierarchical Clustering
  - Difference from k-means
  - Dendrograms
- Gaussian Mixture Models and EM
  - Assumptions
  - Relationship to K-means (soft)
  - Should have working knowledge of fitting a GMM
Feature Selection:

- Difference compared with Dimensionality Reduction
- Filter vs. Wrapper Methods
- Forward Selection
- Backward Selection
- Bi-directional

Dimensionality Reduction:

- PCA
  - Understand meaning of principle components
  - Orthogonal projection onto lower dimensional space
  - Maximize variance / minimize squared error
  - Use for preprocessing
  - Applications
Markov Decision Processes:

- Definition
- What does “Markov” Mean?
- Policies
- MDP search trees
- Reward vs Utility vs Value vs Q function
- Discount factors
- Value Iteration
- Policy Evaluation
- Policy Iteration
Reinforcement Learning:

- Definition, difference compared with MDP problems
- Model-based vs Model-free
- On-policy vs. Off-policy
- Exploration vs. Exploitation
- Q-learning
- Approximate Q-learning
  - With linear Q-functions: relation to least squares regression