Exam Topics

- **Exam #1**
  - machine learning overview
    - features, training data, instances, attributes, test data, class label, target attribute, loss functions
    - supervised, unsupervised, reinforcement
    - regression, classification
    - pros and cons of each algorithm
  - decision tree
    - entropy and information gain (be able to calculate, can use calculator)
    - be able to generate tree
    - categorical features, continuous features, inconclusive leaves
    - rule representation of tree, overfit, underfit, pruning, Occam's razor
  - Bayes optimal classifier
    - representation bias, inductive bias, no free lunch theorem
    - parameters and hyperparameters, train/test/validation
    - reasons classifier may not perform well, accuracy, error
  - geometric algorithms
    - k nearest neighbors, distance function, normalization, inductive bias
    - decision boundaries, boundaries for specific classifiers
    - features as dimensions, one hot encoding
  - k means clustering
    - algorithm, iterative refinement, convergence
    - computational complexity, optimality
    - sum of squared error, elbow method, the curse of dimensionality
    - variations on clustering, density, hierarchical
  - perceptron
    - algorithm, weighted sum of inputs, neuron, activation function, bias term
    - impact of weights, training, modify weights
    - number of iterations, order of training instances, overfit, convergence
    - voting perceptron, averaged perceptron
    - decision boundary, linear separability
    - data preparation, imbalanced class distribution

- **Exam #2**
  - SVMs
    - support vector machines, hard margin, soft margin
    - slack parameters, support vectors, kernels
  - evaluation
    - performance measures, TP, FP, TN, FN, accuracy
    - precision, recall, TP rate, FP rate, PRC, ROC, AUC, confusion matrix
    - macro averaging, micro averaging
    - paired t-test, calculate p-value from t-test, statistical significance
    - confidence intervals, resampling, bootstrapping
  - multi-class classification
    - one vs. all, all vs. all, confusion matrix
  - ranking
    - input, output, preference pairs, bipartite
    - loss functions, Kemeny distance measure
    - weighted inputs, probability distribution over class values
    - naive Rank train/test, ranking classifier with cost function
  - linear regression
    - classification vs. regression
    - gradient descent, learning rate, local minimum
    - mean squared error loss function, derivative
  - regularization
    - relationship between margin and loss, surrogate loss functions
    - role of regularization, types of regularization
    - p-norm, l1 norm, l2 norm, infinity norm
    - lower values of p push toward sparse vectors
    - higher values sensitive to outliers
  - bias
    - bias errors
    - estimation error, approximation error
    - bias and variance, high and low bias and variance
    - regularization and role in bias-variance trade-off
  - naïve Bayes classifier
    - probability, axioms of probability, conditional probability
    - Bayes' rule, naive Bayes assumption and motivation
    - naive Bayes classifier, handle many features, 0 values, continuous values
    - text mining, stop words, tf-idf
  - logistic regression
    - classification, sigmoid function, log loss, loss function convexity
    - return probability of class

- **Exam #3**
  - multi-layer neural networks
    - tanh and sigmoid activation functions and motivations
    - forward propagation to compute output
    - backward propagation to compute weights
    - inductive bias, hyperparameters and impact on underfitting, overfitting
  - deep networks
    - RNN, CNN, autoencoder, variational autoencoder, GAN
    - convolution, softMax, pooling, dropout, ReLU, RNN
  - ensemble classifiers
    - voting, resampling, bagging, boosting, AdaBoost, random forest
    - ways to establish diversity
  - k means++
    - cluster initiation, furthest first, probabilistic selection of cluster means
    - theoretical guarantees
  - dimensionality reduction
    - principal component analysis, principal components
    - projection error, variance, explanation of variance
    - selecting components, skree plot
  - overall
    - remember highlights of algorithms discussed throughout semester
    - be prepared to compare and contrast methods