The first midterm covers in-class material days 1-10, labs 1-5, and reading weeks 1-5. It is in class and closed notes/books/internet/other, but you may use a 1 page (front and back), hand-written "resource sheet" (created by *you*). You will not need a calculator. I have put vocab in blue.

1. Review from CS260
   - Relationship between explanatory variables or features (X) and response variable (y).
   - What is classification? Understand the discrete setting of predicting classes or categories.
   - What is regression? Understand the setting where we predict a continuous response variable.
   - Classification accuracy and relationship to classification error
   - Evaluation metrics for binary classification (confusion matrices, FPR, TPR, ROC curves)
   - Naive Bayes basics including Bayes rule: explain the evidence, prior, posterior, likelihood.
   - How can we predict the label of a new point after fitting a Naive Bayes model?
   - How do we estimate the probabilities of a Naive Bayes model? including Laplace counts
2. Introduction to Machine Learning and ML pipeline
   - How do we define machine learning and why would we want it?
   - Supervised vs. unsupervised learning.
   - Training vs. testing.
   - Common ML notation ($\boldsymbol{X}$, $\boldsymbol{y}$, n, p, etc).
   - What is overfitting? How does it relate to model complexity?
   - Throughout: pros and cons of different ML algorithms.
   - Idea of a loss function, hypothesis space, and generalization error.
   - Sources of error in a pipeline and bias-variance tradeoff
   - Cross-validation and how it can be used for uncertainty and hyper-parameter selection
   - Model cards and intended vs. unintended uses of models
   - Advanced evaluation metrics: AUC, precision/recall curves
   - Most classification methods return *probabilities* and we use a *threshold* for class prediction
   - Throughout: translating ML ideas and algorithms into code (i.e. implementation)
3. K-Nearest Neighbors and KD Trees
   - Understand and use the K-nearest neighbors algorithm (inputs, outputs, conceptual idea).
   - Idea of a distance metric between data points.
   - KD Trees and how to build them using a training dataset.
   - Using a KD tree to find the nearest neighbor(s) of a test point.
   - Runtime of the naive KNN algorithm vs. the KD tree algorithm.

- How the choice of K impacts generalization accuracy.
- How nearest neighbor algorithms can be used in both the classification and regression setting.

4. Decision Trees

- Decision tree as a data structure that can be used for prediction.
- What are the internal nodes of a decision tree? The edges? The leaves?
- What is the depth of a decision tree and how can we choose it to prevent overfitting?
- ID3 decision tree algorithm, use of entropy and conditional entropy to choose best features.
- Conceptual idea of entropy, calculation of entropy
- Different types of stopping criteria when building the tree.

5. Ensemble Methods

- Idea of using an ensemble of classifiers (ideally with low bias) to reduce variance
- To test, let each classifier in the ensemble "vote" (could be weighted or unweighted)
- Bootstrap: sampling from our data with replacement (usually keeping n the same)
- Bagging (Bootstrap Aggregation): create a classifier for each bootstrapped training dataset
- How does averaging the results of many "weak" classifiers reduce the overall error?
- Ensemble notation and example of reducing the error via bagging!
- Random Forest classifiers as ensembles of decision stumps (or small-depth trees)
- What was the idea behind Random Forests? Why might they be better than regular Bagging?
- AdaBoost: upweight training examples that were classified incorrectly in the previous iter
- AdaBoost details: weighted error, score, update example weights, testing with weighted vote
- Decision Trees with weighted examples (how do we modify the probability calculations?)

6. Advanced Regression

- Logistic function of a linear transformation of $\boldsymbol{X}$ as our model in logistic regression.
- Logistic regression creates a *linear* decision boundary (visualize for $p = 1, 2$).
- Idea of a likelihood function and finding the MLE (maximum likelihood estimator).
- Bernoulli random variable example of MLE calculation.
- In logistic regression our cost is the negative log likelihood.
- Derivation of SGD for logistic regression, relationship to linear regression.
- Stochastic gradient descent solution – derivation and implementation details.
- Learning rate α for SGD and how to choose it.
- Idea of multi-class logistic regression and the mathematical details (softmax).
- Adding regularization to gradient descent solutions.
- Regularization based on preventing weights from becoming too large (leads to overfitting).
- Fairness regularization with the goal of creating more equal outcomes for demographic groups.