The first midterm (October 3 in lab + take home) covers in-class material days 1-8, labs 1-4, and reading weeks 1-4. For the in-lab portion (short answer questions), you may bring a 1 page (front and back), hand-written "cheat-sheet" (created by *you*), and a calculator, but no other notes or resources. (You shouldn't need a calculator, but it may make things quicker). I have put vocab in blue.

1. Introduction to Machine Learning
   - How do we define machine learning and why would we want it?
   - How is machine learning similar to and different from related fields?
   - Relationship between explanatory variables or features and response variable.
   - 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.
   - Supervised vs. unsupervised learning.
   - Training vs. testing.
   - Common ML notation ($\boldsymbol{X}$, $\boldsymbol{y}$, n, p, etc).
   - Classification accuracy and relationship to classification error.
   - 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.
2. K-Nearest Neighbors
   - Understand and use the K-nearest neighbors algorithm (inputs, outputs, conceptual idea).
   - Idea of a distance metric between data points.
   - Runtime of K-nearest neighbors and some heuristic ideas for improving it.
   - Interpretation of K-nearest neighbors as a probability (including multi-class prediction).
   - How the choice of K impacts generalization accuracy.
3. 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 (but not Shannon encoding).
   - Different types of stopping criteria when building the tree.
   - How to transform continuous features into binary features? Intuition behind this approach.

4. Linear Regression

- Linear regression problem setup, loss function, error ϵ independent of $\boldsymbol{X}$.
- *Goals* of fitting a linear model to a dataset.
- Squared error loss function in terms of reducible vs. irreducible error.
- MSE (mean squared error) and the general idea of its expected value (not math details).
- Conceptual ideas of bias and variance. What is the bias-variance tradeoff?
- What is a linear function? (notation of $\boldsymbol{w}$ for the weights)
- Goal of minimizing the RSS (residual sum of squared errors) or SSE (sum of squared errors).
- Simple vs. multiple linear regression (+ why do we add a column of 1's?)
- Cost function $J(\boldsymbol{w})$ (add $\frac{1}{2}$ to make derivative work out) and geometric interpretation.
- Closed-form solution (definition and interpretation) for simple and multiple linear regression.
- Stochastic gradient descent solution – derivation and implementation details.
- Learning rate α for SGD and how to choose it.
- Pros and cons of the closed-form solution vs. the SGD solution (vs. batch gradient descent).
- Polynomial regression as an extension of linear regression (for $p = 1$ only).
- *Not* included: adding regularization to both SGD and analytic solutions.

5. Naive Bayes

- Bayes rule and how to apply it; idea of conditional probability.
- Bayes rule in ML: identify and explain the evidence, prior, posterior, likelihood.
- Derivation of the Naive Bayes model for $p(y = k|\vec{x})$ (via the Naive Bayes assumption).
- 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?
- Laplace counts (motivation, application details).
- What types of features/output do we currently require for Naive Bayes?
- Idea of using normalization to compute the evidence (see clinical trials example).
- Confusion matrix as a more nuanced evaluation metric.