The first midterm (February 27 in lab) covers in-class material days 1-10, labs 1-3 + logistic regression, and reading weeks 1-4 (minus Naive Bayes). You may bring a 1 page (front and back), hand-written "cheat-sheet", but no other notes or resources. You will not need a calculator. 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 (see Slides 4 for an example).
   - 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.
- Idea and calculation of expected value (weighted average).
- Expected value of the loss function, reducible vs. irreducible error.
- MSE (mean squared error) and the general idea of its expected value.
- Conceptual ideas of bias and variance. What is the bias-variances tradeoff?
- What is a linear function? (notation of $\boldsymbol{b}$ 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 $J(\boldsymbol{b})$ (add $\frac{1}{2}$ to make derivative work out) and geometric interpretation.
- Analytic solution (derivation, interpretation) for both 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 analytic solution vs. the SGD solution (vs. batch gradient descent).
- Polynomial regression as an extension of linear regression.
- Adding regularization to both SGD and analytic solutions.

5. Logistic Regression

- Why don't we use linear regression for classification problems?
- Logistic function of a linear transformation of $\boldsymbol{X}$ as our model in logistic regression.
- Logistic regression creates a *linear* decision boundary (visualize for $p = 1$).
- 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.
- Intuition behind the cost function.
- Derivation of SGD for logistic regression, relationship to linear regression.
- Idea of multi-class logistic regression (not the mathematical details).