Material on the Midterm & What you need to know:

1. Regression, Penalized Regression, & Regression on Derived Inputs (i.e. PC and PLS Regression).
   - You should understand the properties and comparative strengths and weaknesses of each method.
   - You should know when to apply each method and how to interpret the results.
   - From an applied word problem, you should be able to decide which method would be appropriate and justify your choice.
   - (Stat 640) You should be able to mathematically prove or characterize properties of penalties and penalized regression estimators (i.e. KKT conditions!).

2. Classification: KNN, Nearest Centroid / Naive Bayes, Discriminant Analysis, Logistic / Multinomial Regression, SVMs / Kernel SVMs.
   - You should understand the properties and comparative strengths and weaknesses of each method.
   - You should know when to apply each method and how to interpret the results.
   - From an applied word problem, you should be able to decide which method would be appropriate and justify your choice.
   - (Stat 640) You should be able to mathematically prove or characterize properties of various classifiers.

3. Model Selection & Model Assessment.
   - You should be able to recognize situations where model selection and/or model assessment is done incorrectly or in a way that will bias results.
   - You should be able to set up correct procedures for selecting tuning parameters and assessing the model fit in applied scenarios.

4. Matrix Factorizations: PCA, Sparse PCA, ICA, and NMF.
   - You should understand the properties and comparative strengths and weaknesses of each method.
   - You should know when to apply each method and how to interpret the results.
   - From an applied word problem, you should be able to decide which method would be appropriate and justify your choice.
   - (Stat 640) You should be able to mathematically prove or characterize properties of various matrix factorization techniques.

   - You should understand the properties and comparative strengths and weaknesses of each method.
   - You should know when to apply each method and how to interpret the results.
• From an applied word problem, you should be able to decide which method would be appropriate and justify your choice.

• (Stat 640) You should be able to mathematically prove or characterize properties of various matrix factorization techniques.

6. (Stat 640) Additionally, you should be able to examine a new problem mathematically to understand its properties and relate it to statistical learning methods covered in class.
Sample Midterm Exam Questions

Disclaimer: These are sample questions from past exams that are meant to serve as examples of the types of questions that may be on your midterm. They are not comprehensive and do not reflect the full scope of problems that will be on the exam. Your actual exam questions may be harder and/or easier than these questions.

1. [Stat 640] Let $\hat{\beta}_{\text{ls}}$ be the least squares estimator and $\hat{\beta}_{\text{ridge}}$ be the ridge regression estimator. Prove that $\text{Var}(\hat{\beta}_{\text{ls}}) \geq \text{Var}(\hat{\beta}_{\text{ridge}})$.

2. [Stat 640] Suppose you are fitting a linear regression model with response $Y \in \mathbb{R}^n$ and predictors $X \in \mathbb{R}^{n \times p}$ where the columns of $X$ are orthogonal.
   (a) What is the solution to the lasso problem: minimize $\frac{1}{2}||Y - X\beta||_2^2 + \lambda||\beta||_1$?
   (b) What is the solution to the non-negative lasso problem: minimize $\frac{1}{2}||Y - X\beta||_2^2 + \lambda||\beta||_1$ subject to $\beta_j \geq 0$ for $j = 1, \ldots, p$?

3. [Stat 640 & Stat 444] A marketer is trying to predict a “yes” or “no” response to a survey question based on demographic and social networking data. His training set consists of 74 measured predictors for 826 individuals of which 62 answered “yes” to the survey question. The marketer fits LDA (Linear Discriminant Analysis) and a linear SVM. The training errors for both models are good. When cross-validation is used to assess the misclassification rate, the marketer notices that almost all individuals are classified as “no” and the CV error is much higher than the training error.
   (a) What could be causing this?
   (b) What would you recommend in this situation?

4. [Stat 640 & Stat 444] Instead of using the drop in the proportion of variance explained as shown in the scree plot, some have suggested using cross-validation to select the number of principal components, $K$.
   (a) Outline a procedure to perform 5-fold cross-validation to select $K$, the number of principal components.
   (b) What criteria will you be optimizing to choose $K$?
   (c) How will you leave out data in each fold?
   (d) Will your procedure work well when $n << p$? Explain.

5. [Stat 640 & Stat 444] A business analyst is trying to predict market demand for a product over the next six months. He has 90 features of interest measured from 275 stores
and decides to use the elastic net for his prediction. To select the optimal regularization parameters, he uses five-fold cross-validation. As his boss wants an estimate of the prediction error, he runs five-fold cross-validation again. For each fold, he fits the elastic net with the previously selected regularization parameter value to fourth-fifths of the data and uses the one-fifth left out to estimate the prediction error. He averages the prediction error over each of the five folds and reports this to his boss. Is this an unbiased estimate of the prediction error? If so, why? If not, why not and how would you alter the procedure to obtain an unbiased estimate?

6. [Stat 640] Poisson regression is fit by maximizing the following log-likelihood: \( \ell(\beta) = \sum_{i=1}^{n} \left[ y_i x_i^T \beta - \exp(x_i^T \beta) \right] \). Suppose you want to perform kernel Poisson regression. Write out this log-likelihood in terms of the kernel matrix \( K_{n \times n} \) and parameters \( \alpha_{n \times 1} \).

7. [Stat 640] An eigenvalue problem has the form: maximize \( x^T A x \) subject to \( x^T X = 1 \). Show how to transform Fisher’s Discriminant Analysis into a standard eigenvalue problem.

8. [Stat 640 & Stat 444] For each of the following classification scenarios, which method would you recommend? Why? If you feel like you need more information, specify what information you need and how this information would change your recommendation.
   (a) A scientist cares only about missclassification error. She is trying to predict 10 classes based on 530 samples and 62 predictors.
   (b) A scientist wants to find out which variables are most important for classifying between two classes. He has 180 samples and 5600 features.
   (c) A scientist has data that is highly correlated. She wants to find out which variables are most important for classifying between two classes.
   (d) A scientist wants to classify between two classes with 8000 observations and 64 features. He cares only about prediction error.