Your name: ______________________________

Your SUNet ID: ______________________________

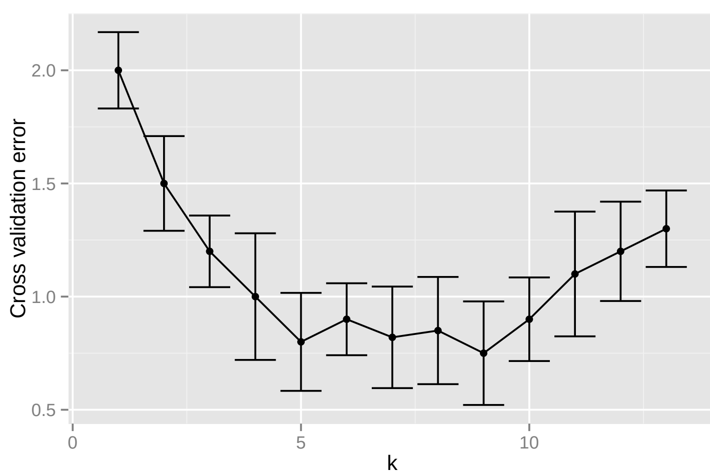
Exam rules:

- You have 75 minutes to complete the exam.
- You are not allowed to consult books or notes, or to use calculator or cell phone. If you must use a computer to type your solutions, you are not allowed to use any software aside from a Word processor or LaTeX.
- A Cheat Sheet is provided at the end of the exam.
- Please show your work and justify your answers.
- **SCPD students:** If you are taking the exam remotely, please return your solutions along with a routing form, signed by your proctor, by 5:50 pm PST on Thursday, July 20. You can email a PDF to scpd-distribution@lists.stanford.edu.

Problem	Points
1	
2	
3	
4	
5	
6	
Total	

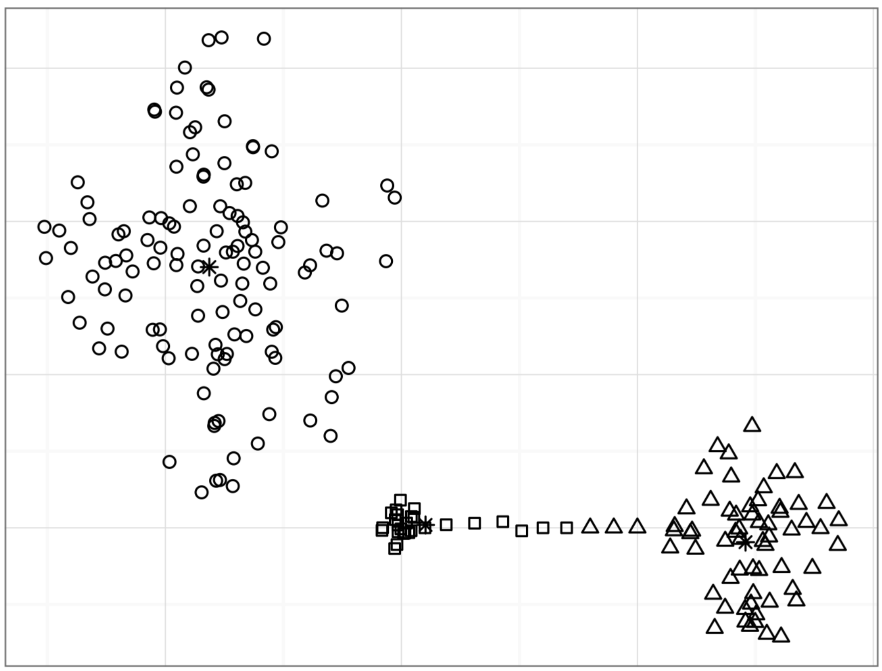
1. **[10 points]** Explain what a *ROC curve* is and how it is used.

2. **[10 points]** State and explain the one standard error rule for model selection using 10-fold cross validation. Apply it to select the optimal number of nearest neighbors in the plot below, which shows the cross-validation error and one standard error intervals as a function of k.



3. **[20 points]** Determine which of the following methods produced the clustering shown below and explain your reasoning. The centroid of each cluster is shown as an asterisk.

   - k-means clustering with $k = 3$.
   - Single linkage hierarchical clustering (dendogram cut at the level where there are 3 clusters).
   - Complete linkage hierarchical clustering (dendogram cut at the level where there are 3 clusters).



4. We define a new kind of discriminant analysis for a classification problem with a binary response. The classes have prior probabilities π_0 and π_1. Given the class, k, the conditional probability of the inputs $X_1, \ldots, X_p$ is multivariate normal with a class-dependent mean μ_k and covariance matrix $\sigma_k \mathbf{\Sigma}$. The matrix $\mathbf{\Sigma}$ is common to both classes and σ_k is a class-dependent constant. All parameters, π_k, μ_k, σ_k, for each class, as well as Σ, are set to their Maximum Likelihood estimates.

   (a) **[10 points]** Provide an equation describing the classifier's decision boundary or discriminant. What would the boundary look like?

   (b) **[10 points]** Why might this classifier be preferable to Linear Discriminant Analysis?

   (c) **[10 points]** Why might this classifier be preferable to Quadratic Discriminant Analysis?

5. Two distances, d and d', are related by a monotone transformation:

$$d'(a, b) = f(d(a, b))$$

which satisfies $f(x) \geq f(y)$ if $x \geq y$.

(a) **[10 points]** Prove that the single linkage hierarchical clustering with k clusters is the same under d and d'.

(b) **[10 points]** Prove that the complete linkage hierarchical clustering with k clusters is the same under d and d'.

6. **[10 points]** Let $(X_1, Y_1), ..., (X_n, Y_n) \overset{iid}{\sim} P_0$ and assume we fit a logistic regression to our data with no intercept. Specifically, we fit the following model

$$\log \left[\frac{\mathbb{P}[Y = 1|\mathbf{X}]}{\mathbb{P}[Y = 0|\mathbf{X}]} \right] = \beta_1 X_1, \tag{1}$$

Derive the gradient for β_1.

Cheat sheet

The sample variance of $x_1, \dots, x_n$ is:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

The residual sum of squares for a regression model is:

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

t-test:

The t-statistic for hypothesis $H_0 : \beta_i = 0$ is

$$t = \frac{\hat{\beta}_i}{\mathrm{SE}(\hat{\beta}_i)}$$

F-test:

The F-statistic for hypothesis $H_0 : \beta_{p-q+1} = \beta_{p-q+2} = \cdots = \beta_p = 0$ is

$$F = \frac{(RSS_0 - RSS)/q}{RSS/(n-p-1)},$$

where RSS_0 is the residual sum of squares for the null model H_0, and RSS is the residual sum of squares for the full model with all predictors. Asymptotically, the F-statistic has the F-distribution with degrees of freedom $d_1 = q$ and $d_2 = n - p - 1$.

Minimum F-statistic to reject H_0 at a significance level $\alpha = 0.01$

		d_1			
		1	2	3	4
	1	4052.181	4999.500	5403.352	5624.583
	10	10.044	7.559	6.552	5.994
d_2	20	8.096	5.849	4.938	4.431
	30	7.562	5.390	4.510	4.018
	120	6.851	4.787	3.949	3.480

Logistic regression:

Logistic regression assigns to positive if the estimated conditional probability

$$\hat{P}(Y = + | X = x) = \frac{e^{X\cdot\hat{\beta}}}{1 + e^{X\cdot\hat{\beta}}}$$

LDA:

The log-posterior of class k given an input x is:

$$C + \log \pi_k - \frac{1}{2}\mu_k^T \mathbf{\Sigma}^{-1}\mu_k + x^T \mathbf{\Sigma}^{-1}\mu_k$$

where C is a constant which does not depend on k.

QDA:

The log-posterior of class k given an input x in QDA is:

$$C + \log \pi_k - \frac{1}{2}\mu_k^T \mathbf{\Sigma}_k^{-1}\mu_k + x^T \mathbf{\Sigma}_k^{-1}\mu_k - \frac{1}{2}x^T \mathbf{\Sigma}_k^{-1} x - \frac{1}{2}\log|\mathbf{\Sigma}_k|$$

where C is a constant which does not depend on k.