STAT 6910-003 – SLDM II – Homework #2

Due: 5:00 PM 10/26/18

1. **Linear Algebra Review (10 pts)**

   (a) (5 pts) Show that if U is an orthogonal matrix, then for all $\mathbf{x} \in \mathbb{R}^d$, $\|\mathbf{x}\| = \|U\mathbf{x}\|$, where $\|\cdot\|$ indicates the Euclidean norm.

   (b) (5 pts) Show that all 2×2 orthogonal matrices have the form

   $$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}.$$

   Give a geometric interpretation of the effect of these two transformations.

2. **Probability (18 pts)**

   (a) (9 pts) Let random variables X and Y be jointly continuous with pdf $p(x, y)$. Prove the following results:

      i. $\mathbb{E}[X] = \mathbb{E}_Y[\mathbb{E}_X[X|Y]]$ where $\mathbb{E}_Y$ is the expectation with respect to Y

      ii. $\mathbb{E}[\mathbf{1}[X \in C]] = \Pr(X \in C)$, where $\mathbf{1}[X \in C]$ is the indicator function of an arbitrary set C. That is, $\mathbf{1}[X \in C] = 1$ if $X \in C$ and 0 otherwise.

      iii. If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

   (b) (9 pts) For the following equations, describe the relationship between them. Write one of four answers to replace the question mark: "$=$", "$\leq$", "$\geq$", or "depends". Choose the most specific relation that always holds and briefly explain why. Assume all probabilities are non-zero.

      i. $\Pr(X = x, Y = y) \,?\, \Pr(X = x)$

      ii. $\Pr(X = x|Y = y) \,?\, \Pr(X = x)$

      iii. $\Pr(X = x|Y = y) \,?\, \Pr(Y = y|X = x)\Pr(X = x)$

3. **Positive (semi-)definite matrices (25 pts)**. Let A be a real, symmetric $d \times d$ matrix. We say A is positive semi-definite (PSD) if for all $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{x}^T A\mathbf{x} \geq 0$. A is positive definite (PD) if for all $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^T A\mathbf{x} > 0$. We write $A \succeq 0$ and $A \succ 0$ when A is PSD or PD, respectively.
   The spectral theorem says that every real symmetric matrix A can be expressed via the spectral decomposition

   $$A = U\Lambda U^T$$

   where U is a $d \times d$ orthogonal matrix and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$.
   Using the spectral decomposition, show that

   (a) (5 pts) If $\mathbf{u}_i$ is the i-th column of U then $\mathbf{u}_i$ is an eigenvector of A with corresponding eigenvalue λ_i.

   (b) (10 pts) A is PSD iff $\lambda_i \geq 0$ for each i.

   (c) (10 pts) A is PD iff $\lambda_i > 0$ for each i.
   Hint: For parts (b) and (c), use the identity

   $$U\Lambda U^T = \sum_{i=1}^{d} \lambda_i \mathbf{u}_i \mathbf{u}_i^T,$$

   which can be verified by showing the matrices on both sides of the equation have the same entries.

4. **The Bayes Classifier (37 pts).** Let X be a random variable representing a 1-dimensional feature space and let Y be a discrete random variable taking values in $\{0, 1\}$ (i.e., Y is the corresponding class label). If $Y = 0$, then the posterior distribution of X for class 0 is Gaussian with mean μ_0 and variance σ_0^2. If $Y = 1$, then the posterior distribution of X for class 1 is Gaussian with mean μ_1 and variance σ_1^2. Let $w_0 = \Pr(Y = 0)$ and $w_1 = \Pr(Y = 1) = 1 - w_0$.

   (a) (5 pts) Derive the Bayes classifier for this problem as a function of w_i, μ_i, and σ_i where $i \in \{0, 1\}$.

   (b) (5 pts) Derive the Bayes error rate for this classification problem as a function of w_i, μ_i, and σ_i where $i \in \{0, 1\}$. You may write your solution in terms of the Q function where if Z is a standard normal random variable, then $Q(z) = \Pr(Z > z)$.

   (c) (5 pts) Describe how to perform cross-validation for a classification problem.

   (d) Set $\mu_0 = 0$, $\mu_1 = 1.5$, $\sigma_0 = \sigma_1 = \sigma = 1$, $w_0 = 0.3$ and $w_1 = 0.7$. For the sample sizes $N \in \{100, 200, 500, 1000\}$, simulate the above classification problem. Apply the Bayes classifier, logistic regression, and the k-nearest neighbor (nn) classifier to the simulated data. Run this simulation for 100 trials and report the following for each sample size:

      i. (3 pts) The Bayes error rate.

      ii. (3 pts) The average value of k as selected using cross-validation.

      iii. The classification error of each classifier. Calculate the error of the logistic regression and k-nn classifiers for each trial using cross-validation and describe how you performed cross-validation (e.g. 5-fold, 10-fold cross validation, etc.; 1 pt). You may use built-in logistic regression and k-nn classifier functions (i.e. you do not need to code these up from scratch). Report the mean and standard deviation of the error in

         A. (6 pts) Table form

         B. (9 pts) Graphical form. Make a plot with sample size on the x-axis and the error on the y-axis. Plot the mean and standard deviation using error bars. Plot the results for all 3 classifiers on the same plot. You may need to use log scales for better visualization.

         You do not need to turn in your code.

5. How long did this assignment take you? (5 pts)

6. Type up homework solutions (5 pts)