Homework I

STAT 6655 - Spring semester 2024

Due: Friday, January 26, 2024 - <u>5:00 PM</u>

Please upload your solutions in a single pdf file in Canvas. Any requested plots should be sufficiently labeled for full points. Include any code requested.

Unless otherwise stated, programming assignments should use built-in functions in your chosen programming language (Python, R, or Matlab). However, exercises are designed to emphasize the nuances of machine learning algorithms - if a function exists that trivially solves an entire problem, please consult with the TA before using it.

Please be clear in your explanations and provide the necessary steps to reach the solution. You are not required to type the solutions, but unorganized work and difficult to read handwriting may be subject to point deductions to the TA discretion.

1. **Two cultures (10 pts)**. Read the paper by Leo Breiman about two cultures within the statistics community (found under "Files/Papers" on Canvas). Write a short summary of the main points in the paper. Are there any points you particularly agree or disagree with? Your total response should be approximately 1-2 paragraphs.

2. **Probability**

   (a) (**15 pts**) Let random variables X and Y be jointly continuous with pdf $p(x, y)$. Prove the following results:

   i. (**5810 - 5, 6655 - 3**) If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$

   ii. (**5810 - 5, 6655 - 3**) $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

   iii. (**5810 - 5, 6655 - 3**) If X and Y are independent, then $\text{var}[X + Y] = \text{var}[X] + \text{var}[Y]$

   iv. (**6655 - 3**) $\mathbb{E}[X] = \mathbb{E}_Y[\mathbb{E}_X[X|Y]]$ where $\mathbb{E}_Y$ is the expectation with respect to Y

   v. (**6655 - 3**) $\mathbb{E}[\mathbf{1}[X \in C]] = \Pr(X \in C)$, 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.

(b) (**6655 - 5 pts**) If a random variable can take two possible values $x \in \{-1, 1\}$, show that the following is a valid probability mass function for an appropriate choice of u:

$$p(x|u) = \left(\frac{1-u}{2}\right)^{(1-x)/2} \left(\frac{1+u}{2}\right)^{(1+x)/2}.$$

Find the interval of valid values for u, $\mathbb{E}[x]$, and $\text{var}[x]$.

3. **Linear Algebra Review** .

Positive (semi-)definite matrices. 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 $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)$.

(a) (**6655 - 8 pts**) Using the spectral decomposition, show that A is PSD iff $\lambda_i \geq 0$ for each i. *Hint:* 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 (you do not have to verify this).

(b) (**5810 - 6 pts, 6655 - 4 pts**) Using the spectral decomposition, show that 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.

(c) (**5810 - 6 pts, 6655 - 4 pts**) Using the definition of a PD matrix, prove that the sum of two PD matrices is also PD. A very similar approach can be used to prove the sum of two PSD matrices is also PSD (although you don't have to prove it).

(d) (**5810 - 3 pts**) Is the sum of a PD matrix and a PSD matrix necessarily PD, PSD, or neither? Explain why.

(e) (**5810 - 8 pts**) Consider the following matrices:

$$A = \begin{bmatrix} 1 & -2 \\ 3 & 4 \\ -5 & 6 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, C = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}$$

Determine whether the following matrices are PD, PSD, or neither. Briefly explain why. You may use the NumPy library for this problem.

i. A

ii. $A^T A$

iii. AA^T

iv. B

v. $-B$

vi. C

vii. $C - 0.1 \times B$

viii. $C - 0.01 \times AA^T$

(f) (**5810 - 7 pts, 6655 - 4 pts**) Recall that $U \in \mathbb{R}^{d \times d}$ is an orthogonal matrix if its transpose is its inverse, i.e., $U^T U = UU^T = I$. 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. *Hint:* Use the relationship between the squared Euclidean norm and the dot product, i.e. $\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$.

4. **Multivariate calculus (20 pts)**.

(a) (**5810**) Consider the function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ as defined below:

$$\mathbf{f}(\mathbf{v}) = \begin{bmatrix} f_1(\mathbf{v}) \\ f_2(\mathbf{v}) \end{bmatrix} = \begin{bmatrix} v_1^2 v_2 + 3v_1 e^{v_2^2} \\ 4v_1^3 v_2 - v_1 \log(v_2) \end{bmatrix}.$$

i. Compute the gradient and Hessian of f_1.

ii. Compute the gradient and Hessian of f_2.

iii. Compute the Jacobian of $\mathbf{f}$.

(b) (**6655**) Consider the following function:

$$l(u, \Sigma; x_1, x_2, \ldots, x_n) = \log \left(\prod_i^n \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x_i - u)^T \Sigma^{-1} (x_i - u) \right] \right),$$

where $x_i, u \in \mathbb{R}^p$, and $\Sigma \in \mathbb{R}^{p \times p}$ is a positive definite symmetric matrix. Compute the gradient of l with respect to u, set it equal to the zero vector, and then solve for u.

5. **Estimation** (**5810 - 10 pts**, **6655** - 15 pts). Suppose we have random variables $X_1, \ldots, X_n$ that are independent and identically distributed each with mean μ and variance σ^2. In this setting, the variables X_i can be viewed as data points sampled in such a way that we expect each data point to be drawn from the same distribution. In general, we do not know this distribution including its properties such as the mean and variance. Thus if we want to know these properties, we need to estimate them from the data.

Suppose that we wish to estimate some parameter c with an arbitrary estimator $\hat{c}$. Since the estimator depends on the data, it is also a random variable. The ***bias*** of this estimator is equal to $\mathbb{E}[\hat{c}] - c$. Define the following:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

$\bar{X}$ is an estimator of μ.

(a) (3 pts) Calculate the bias of $\bar{X}$.

(b) (3 pts) Calculate the variance of the estimator $\bar{X}$.

Hint: The variance of the sum of INDEPENDENT random variables is equal to the sum of the variances. Also, if a is a constant scalar, then $Var[aX] = a^2 Var[X]$.

(c) (4 pts) The mean squared error (MSE) of an estimator is defined as

$$MSE(\hat{c}) = \mathbb{E}[(\hat{c} - c)^2].$$

It can be shown that the MSE of an estimator is equal to the square of its bias plus the variance. What is the MSE of the estimator $\bar{X}$?

(d) (**6655** - 5 pts) Consider the following estimator of σ^2:

$$\hat{s}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Calculate the bias of $\hat{s}^2$. If this estimator is biased, is there a way to define a new estimator of σ^2 that is unbiased?

6. **KNN Classifier (15 pts)** Load the dataset from file assignment1.zip. You can find it under the Data folder in the Canvas page. Apply a k-nn classifier to the data with $k \in \{1, 5, 10, 15\}$. Calculate the test error using an approach called 5-fold cross validation (built-in classifiers should have an option for this). Plot the test error as a function of k. Which value of k performs best? You may use any built-in k-nn classifiers in Python, R, or Matlab.