STAT 6910-001 – Principles of ML – Homework #1

Due: 5:00 PM 9/13/19

1. **Two cultures (19 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. **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.

3. **Probability (15 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) (6 pts) For the following equations, describe the relationship between them. Write one of four answers to replace the question mark: "=", "≤", "≥", 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. $\Pr(X = x|Y = y) \,?\, \Pr(Y = y|X = x)\Pr(X = x)$

4. **Positive (semi-)definite matrices (36 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) (8 pts) A is PSD iff $\lambda_i \geq 0$ for each i.

(c) (8 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.

(d) (5 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).

(e) (2 pts) Is the sum of a PD matrix and a PSD matrix necessarily PD, PSD, or neither? Explain why.

(f) (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 any numerical software 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. $C - 0.01 \times AA^T$

5. **Multivariate calculus (10 pts)**. Consider the function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ as defined below:

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

(a) (4 pts) Compute the gradient and Hessian of f_1.

(b) (4 pts) Compute the gradient and Hessian of f_2.

(c) (2 pts) Compute the Jacobian of $\mathbf{f}$.

6. **Estimation (10 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) (2 pts) Calculate the bias of $\bar{X}$.

(b) (2 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) (2 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) (4 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?