Boston University
Department of Electrical and Computer Engineering

ENG EC 414 Introduction to Machine Learning

HW 1

Issued: Wed 2 Sept 2020 **Due:** 4:55pm Tue 8 Sept 2020 on Gradescope

---

Important: Before you proceed, please read the documents pertaining to *Homework formatting and submission guidelines* in the Homeworks section of Blackboard.

Note: Problem difficulty = number of coffee cups ☕.

Problem 1.1 *(Linear Algebra)* [22pts] Let $\mathbf{v}_1 = [1,1,0]^\top$, $\mathbf{v}_2 = [0,1,1]^\top$, and $\mathbf{v}_3 = [1,1,1]^\top$, be three column vectors. Note: $^\top$ means transpose.

(a) [1pt] The dimension d of the Euclidean space $\mathbb{R}^d$ containing $\mathbf{v}_1$ is:

(b) [1pt] The length, i.e., norm $\|\mathbf{v}_1\|$, of $\mathbf{v}_1$ is:

(c) [1pt] The dot product, i.e., inner product $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \mathbf{v}_2^\top \mathbf{v}_1$, of $\mathbf{v}_1$ and $\mathbf{v}_2$ is:

(d) [1pt] Are $\mathbf{v}_1$ and $\mathbf{v}_2$ perpendicular (orthogonal)? Yes/No, Why?

(e) [1pt] Are $\mathbf{v}_1$ and $\mathbf{v}_2$ linearly independent? Yes/No, Why?

(f) [5pts] ☕ If $\mathrm{Proj}_{\mathcal{S}}(\mathbf{v}_3) = a_1\mathbf{v}_1 + a_2\mathbf{v}_2$, where a_1, a_2 are scalars, denotes the orthogonal projection of $\mathbf{v}_3$ onto the subspace $\mathcal{S}$ spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$, then $\mathbf{a} = (a_1, a_2)^\top =$

(g) [5pts] ☕☕ Consider the following subset of $\mathbb{R}^3$: $\mathcal{S} := \{\mathbf{x} : \mathbf{v}_3^\top \mathbf{x} - 3 = 0\}$. Compute the Euclidean distance of $\mathbf{v}_1$ from $\mathcal{S}$ and the point in $\mathcal{S}$ which is closest to $\mathbf{v}_1$.

(h) [4pts] Let $B = \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}$. Compute: (i) its eigenvalues and (ii) a set of orthonormal eigenvectors.

(i) [3pts] The trace $\mathrm{tr}(D)$ of a square matrix D is the sum of all its elements along the main diagonal. Let $D = ABC$, where the dimensions of A, B, and C are, respectively, $p \times q$, $q \times r$, and $r \times p$. What is the relationship between: $\mathrm{tr}(ABC)$, $\mathrm{tr}(BCA)$, and $\mathrm{tr}(CAB)$? Explain.

Problem 1.2 *(Multivariate Calculus)* [9pts] Let A be a $d \times d$ matrix and $\mathbf{b}, \mathbf{x} \in \mathbb{R}^d$ be two $d \times 1$ column vectors. Let $f(\mathbf{x})$ denote a real-valued function of d variables (d components of $\mathbf{x}$).

(a) [2pts] Compute the gradient vector $\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \ldots, \frac{\partial f}{\partial x_d}(\mathbf{x})\right)^\top$ when $f(\mathbf{x}) = \mathbf{b}^\top \mathbf{x}$.

(b) [3pts] Compute the gradient vector $\nabla f(\mathbf{x})$ when $f(\mathbf{x}) = \mathbf{x}^\top A \mathbf{x}$.

(c) [4pts] ☕ Let A be symmetric and invertible. If $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A \mathbf{x} + \mathbf{b}^\top \mathbf{x}$, then find $\mathbf{x}$'s for which $f(\mathbf{x})$ is minimum or maximum.

Problem 1.3 *(Probability)* [21pts] Let $W = Y + U$, where label Y and additive noise U are independent random variables, $P(Y = +1) = 3/4$, $P(Y = -1) = 1/4$, and U is continuous with $U \sim \text{Uniform}[-3, 3]$.

Let observation (feature) $X = 2 \times 1(W > 0) - 1$. where $1(\textit{event})$ is the indicator function of the *event* and equals one if the *event* is true and equals zero if the *event* is false.

(a) [2pts] Sketch the graph of $p(w|Y = +1)$. Proper labeling of axes and key points is needed to receive full credit.

(b) [2pts] Compute the joint pmf $p(x, y) = P(X = x, Y = y)$ for all $x, y \in \{-1, +1\}$.

(c) [1pt] Compute the marginal pmf $p(x) = P(X = x)$, for $x = -1, 1$.

(d) [2pts] Compute the mean/expectation: $\mu_X = E[X]$ and $\mu_Y = E[Y]$.

(e) [2pts] Compute the variance: $\sigma_X^2 = \text{var}(X)$.

(f) [2pts] Compute the correlation $E[XY]$. Are X and Y orthogonal? Explanation is needed to receive credit.

(g) [2pts] Compute the covariance: $\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$. Are X and Y uncorrelated? Explanation is needed to receive credit.

(h) [2pts] Compute the conditional pmf $p(x|y) = P(X = x|Y = y)$ for all $x, y \in= \{-1, +1\}$.

(i) [2pts] Are X and Y independent? Explanation is needed to receive credit.

(j) [4pts] Compute the conditional expectation/mean $E[Y|X = x]$ as a function of x.

Problem 1.4 *(Random Vectors)* [10pts] Let Z_1, Z_2 be i.i.d. (scalar) standard Gaussians (normals) $\mathcal{N}(0, 1)$, i.e., independent and identically Gaussians with zero mean and unit variance. Let

$$\underbrace{\begin{bmatrix} X \\ Y \end{bmatrix}}_{\mathbf{U}} = \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}}_{\mathbf{Z}} + \underbrace{\begin{bmatrix} -3 \\ 2 \end{bmatrix}}_{\mathbf{v}}$$

(a) [2pts] Compute the 2×1 mean vector $E[\mathbf{U}]$.

(b) [4pts] Compute the 2×2 (auto- or self-) covariance matrix $\text{Cov}(\mathbf{U})$ of the random vector $\mathbf{U} = (X, Y)^\top$.

(c) [4pts] Compute the 2×2 cross-covariance matrix $\text{Cov}(\mathbf{U}, \mathbf{Z})$ of the random vector $\mathbf{U}$ and the random vector $\mathbf{Z} = (Z_1, Z_2)^\top$.