Boston University
Department of Electrical and Computer Engineering

ENG EC 414 Introduction to Machine Learning

HW 1 Solution

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:

Solution: 3

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

Solution: $\sqrt{1^2 + 1^2 + 0^2} = \sqrt{2}$.

(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:

Solution: $1 \cdot 0 + 1 \cdot 1 + 0 \cdot 1 = 1$.

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

Solution:

No, because their inner product is not zero.

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

Solution:

Yes: Let $V = [\mathbf{v}_1, \mathbf{v}_2]$ and $\boldsymbol{a} = [a_1, a_2]^\top$. The only solution to the equation $V\mathbf{a} = 0$, which tests the linear dependence of $\mathbf{v}_1$ and $\mathbf{v}_2$, is $\mathbf{a} = (V^\top V)^{-1}\mathbf{0} = \mathbf{0}$. Note: $(V^\top V)$ is an invertible matrix (its determinant is not zero).

(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 $\mathbf{a} = (a_1, a_2)^\top =$

Solution: $(2/3, 2/3)^\top$:

Let $V = [\mathbf{v}_1, \mathbf{v}_2]$. Then $\mathrm{Proj}_{\mathcal{S}}(\mathbf{v}_3) = V\mathbf{a}$. According to the orthogonality principle, $\mathbf{v}_3 - \mathrm{Proj}_{\mathcal{S}}(\mathbf{v}_3)$ is orthogonal to all vectors in $\mathcal{S}$, in particular, both $\mathbf{v}_1$ and $\mathbf{v}_2$. Thus, $V^\top(\mathbf{v}_3 - V\mathbf{a}) = 0$. Solving, we get $\mathbf{a} = (V^\top V)^{-1} V^\top \mathbf{v}_3 = (2/3, 2/3)^\top$, and $\mathrm{Proj}_{\mathcal{S}}(\mathbf{v}_3) = (2/3, 4/3, 2/3)^\top$.

(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 $\mathbf{v}_1$.

Solution:

The point

$$\mathbf{v} := \mathbf{v}_1 - (\mathbf{v}_3^\top \mathbf{v}_1 - 3)\frac{\mathbf{v}_3}{\|\mathbf{v}_3\|^2} = (1\tfrac{1}{3}, 1\tfrac{1}{3}, \tfrac{1}{3})^\top$$

lies in $\mathcal{S}$ since $\mathbf{v}_3^\top \mathbf{v} - 3 = 0$. If $\mathbf{u}$ is any other point in $\mathcal{S}$, then we have $\mathbf{v}_3^\top \mathbf{u} - 3 = 0$ and therefore $\mathbf{v}_3^\top(\mathbf{v} - \mathbf{u}) = 0$. Thus,

$$\|\mathbf{v}_1 - \mathbf{u}\|^2 = \|(\mathbf{v}_1 - \mathbf{v}) + (\mathbf{v} - \mathbf{u})\|^2 = \|\mathbf{v}_1 - \mathbf{v}\|^2 + \|\mathbf{v} - \mathbf{u}\|^2 + 2\underbrace{(\mathbf{v}_1 - \mathbf{v})^\top}_{\propto \mathbf{v}_3^\top}\cancelto{0}{(\mathbf{v} - \mathbf{u})} \geq \|\mathbf{v}_1 - \mathbf{v}\|^2.$$

This shows that $\mathbf{v}$ is the point in $\mathcal{S}$ that is closest to $\mathbf{v}_1$. The minimum distance is equal to $\|\mathbf{v}_1 - \mathbf{v}\| = |\mathbf{v}_3^\top \mathbf{v}_1 - 3|/\|\mathbf{v}_3\| = 1/\sqrt{3}$.

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

Solution:

(i) Eigenvalues: $\lambda = 1, 7$ obtained as the solutions to the quadratic equation $\det(B - \lambda I) = 0$ where I is the 2×2 identity matrix. (ii) One set of orthonormal eigenvectors (not unique): $\mathbf{u}_1 = \frac{1}{\sqrt{2}}(1, 1)^\top$ and $\mathbf{u}_2 = \frac{1}{\sqrt{2}}(1, -1)^\top$. Note: orthonormal means, mutually orthogonal (zero inner product) and unit norm (length).

(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.

Solution:

They are all equal! Proof: $\mathrm{tr}(D) = \sum_i D_{ii}$. Also, $D_{ij} = \sum_{k,l} A_{ik} B_{kl} C_{lj}$. Thus, $\mathrm{tr}(D) = \sum_{i,k,l} A_{ik} B_{kl} C_{li}$ which is symmetric with respect to circular re-orderings of $A - B - C$ (in that order).

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 $f(\mathbf{x}) = \mathbf{b}^\top \mathbf{x}$.

Solution:

$f(\mathbf{x}) = \sum_i b_i x_i \Rightarrow \frac{\partial f}{\partial x_i}(\mathbf{x}) = b_i \Rightarrow \nabla f(\mathbf{x}) = \mathbf{b}$.

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

Solution:

Let $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{b}$ where $\mathbf{b} = A\mathbf{x}$ is a function of $\mathbf{x}$. Then $f(\mathbf{x}) = \sum_j x_j b_j$. By the chain rule,

$$\frac{\partial f}{\partial x_i} = \sum_j \left[b_j \frac{\partial x_j}{\partial x_i} + x_j \frac{\partial b_j}{\partial x_i} \right] = b_i + \sum_j x_j A_{ji} = \sum_j (A_{ij} + A_{ji}) x_j = \sum_j (A + A^\top)_{ij} x_j.$$

Hence, $\nabla f(\mathbf{x}) = (A + A^\top)\mathbf{x}$.

(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.

Solution:

$f(\mathbf{x})$ is differentiable everywhere and is therefore minimized or maximized when $\nabla f(\mathbf{x}) = 0$ or at infinity. Using parts (a) and (b) and the fact that $A = A^\top$ (symmetric matrix) we get $\nabla f(\mathbf{x}) = A\mathbf{x} + \mathbf{b} = 0 \Rightarrow \mathbf{x} = -A^{-1}\mathbf{b}$ is the unique solution (since A is invertible). Thus $f(\mathbf{x})$ is minimum or maximum at $\mathbf{x} = -A^{-1}\mathbf{b}$ or at infinity. If A was positive definite, then $f(\mathbf{x})$ would be minimum at $\mathbf{x} = -A^{-1}\mathbf{b}$ and maximum (in fact, $+\infty$) at infinity. If A was negative definite, then $f(\mathbf{x})$ would be maximum at $\mathbf{x} = -A^{-1}\mathbf{b}$ and minimum (in fact, $-\infty$) at infinity. If A was neither positive nor negative definite, then the maximum is $+\infty$ and the minimum is $-\infty$ and both occur at infinity (along different directions).

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.

Solution: The graph of $p(w|Y = +1)$ is piecewise constant. It is equal to the constant $1/6$ from $w = -2$ to $w = 4$ and the constant 0 for all other values of w. Key break points to indicate: $w = -2, 4$. Key values to indicate $1/6$. Label w for horizontal axis.

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

Solution: *Note:* $2 \times 1(W > 0) - 1 = \text{sign}(W)$. Thus both X and Y are discrete and take only the two values: ± 1 with nonzero probability. $P(X = +1, Y = +1) = P(W > 0, Y = +1) = P(W > 0|Y = +1) \times P(Y = +1) = 4/6 \times 3/4 = 1/2$. Similarly, for other x, y:

$p(x, y) =$

	$x = -1$	$x = +1$
$y = +1$	$1/4$	$1/2$
$y = -1$	$1/6$	$1/12$

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

Solution: $P(X = -1) = P(X = -1, Y = +1) + P(X = -1, Y = -1) = 1/4 + 1/6 = 5/12$. Also, $P(X = +1) = 1 - P(X = -1) = 7/12$

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

Solution: We have, $\mu_X = E[X] = -1 \times 5/12 + 1 \times 7/12 = 1/6$. Similarly, $\mu_Y = E[Y] = -1 \times 1/4 + 1 \times 3/4 = 1/2$.

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

Solution:

We have $\sigma_X^2 = \text{var}(X) = E[X^2] - (E[X])^2 = 1 - (1/6)^2 = 35/36$.

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

Solution:

$E[XY] = 1 \times P(X = 1, Y = 1) + 1 \times P(X = -1, Y = -1) - 1 \times P(X = -1, Y = +1) - 1 \times P(X = +1, Y = -1) = 1/2 + 1/6 - 1/4 - 1/12 = 1/3$. No, X and Y are not orthogonal because their correlation is not equal to zero.

(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.

Solution: $\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y = 1/3 - 1/6 \times 1/2 = 1/4 \neq 0$. No, X and Y are not uncorrelated because $\text{cov}(X, Y) \neq 0$.

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

Solution: *Note:* both X and Y are discrete and take only the two values: ± 1 with nonzero probability. $P(X = +1|Y = +1) = P(W > 0|Y = +1) = 4/6 = 2/3$. Similarly, for other x, y:

$p(x|y) =$

	$x = -1$	$x = +1$
$y = +1$	$1/3$	$2/3$
$y = -1$	$2/3$	$1/3$

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

Solution: X, Y not independent: $P(X = +1|Y = +1) = 2/3 \neq P(X = +1|Y = -1) = 1/3$.

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

Solution: $E[Y|X = x] = 1 \cdot P(Y = +1|X = x) - 1 \cdot P(Y = -1|X = x) = 2P(Y = 1|X = x) - 1 = 2P(X = x|Y = 1) \cdot P(Y = 1)/P(X = x) - 1 = \begin{cases} 2 \cdot \frac{2/3 \cdot 3/4}{2/3 \cdot 3/4 + 1/3 \cdot 1/4} - 1 = 5/7 & \text{if } x = +1, \\ 2 \cdot \frac{1/3 \cdot 3/4}{1/3 \cdot 3/4 + 2/3 \cdot 1/4} - 1 = 1/5 & \text{if } x = -1. \end{cases}$

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 $\mathbf{Z} = (Z_1, Z_2)^\top$.

Solution:

(a) [2pts] Since Z_1 and Z_2 have zero means, we have $E[\mathbf{U}] = A\,E[\mathbf{Z}] + \mathbf{v} = \mathbf{v} = (-3, 2)^\top$.

(b) [4pts] Since Z_1 and Z_2 have zero means, are independent (therefore uncorrelated), and have and unit variances, $\text{Cov}(\mathbf{Z}) = I_2$, the 2×2 identity matrix. Thus, $\text{Cov}(\mathbf{U}) = A \cdot \text{Cov}(\mathbf{Z}) \cdot A^\top = A \cdot I_2 \cdot A^\top = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$.

(c) [4pts] $\text{Cov}(\mathbf{U}, \mathbf{Z}) = \text{Cov}(A\mathbf{Z} + \mathbf{v}, \mathbf{Z}) = A \cdot \text{Cov}(\mathbf{Z}) = A \cdot I_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.