EC503 Introduction to Learning from Data — Boston University
Francesco Orabona — Spring 2021

Homework 1
Due: Tuesday, February 9, 3:30pm

Linear Algebra Problems

Problem 1.1 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, with elements as listed.

(a) Are $\mathbf{v}_1$ and $\mathbf{v}_2$ orthogonal? Yes/No, Why?

(b) Are $\mathbf{v}_1$ and $\mathbf{v}_2$ linearly independent? Yes/No, Why?

(c) Let $\text{Proj}_{\mathcal{S}}(\mathbf{v}_3) = a_1\mathbf{v}_1 + a_2\mathbf{v}_2$, where a_1, a_2 are scalars, denote the orthogonal projection of $\mathbf{v}_3$ onto the subspace $\mathcal{S}$ spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$. Find $\mathbf{a} = [a_1, a_2]^\top$ and $\text{Proj}_{\mathcal{S}}(\mathbf{v}_3)$.

(d) Let $\mathcal{S} := \{\mathbf{x} : \mathbf{v}_3^\top \mathbf{x} = 0\}$. Compute the Euclidean distance of $\mathbf{v}_1$ from $\mathcal{S}$, that is, the minimum distance $\|\mathbf{v}_1 - \mathbf{x}\|_2$ over any element $\mathbf{x}$ of $\mathcal{S}$.

(e) The trace $\text{tr}(D)$ 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: $\text{tr}(ABC)$, $\text{tr}(BCA)$, and $\text{tr}(CAB)$? Explain.

Problem 1.2

(a) Prove that $A \in \mathbb{R}^{n\times n}$ and $A^\top$ have the same eigenvalues.

(b) Let λ_i are the eigenvalues of $M \in \mathbb{R}^{n\times n}$. Determine the eigenvalues of $\alpha M + \beta I$, where I is the identity matrix, and $\alpha, \beta \in \mathbb{R}$.

Multivariate Calculus

Problem 1.3

(a) Let $f(\mathbf{x}) = 2\mathbf{x}^\top A\mathbf{x} + \mathbf{x}^\top \mathbf{b} + c$, where $\mathbf{x} \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$ symmetric, $b \in \mathbb{R}^d$, and $c \in \mathbb{R}$. Find the gradient.

(b) Let $f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|_2^2$, where $\mathbf{x} \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$, $b \in \mathbb{R}^d$. Find the gradient.

(c) Let $f(\mathbf{x}) = \|\mathbf{x}\|_2$ where $\mathbf{x} \in \mathbb{R}^d$. Find the gradient in any $\mathbf{x} \neq \mathbf{0}$. Hint: use directly the definition of gradient with the partial derivatives.

Probability Problems

Problem 1.4 The Gaussian distribution of parameters μ and σ^2 had pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} .$$

Prove that the mean and the variance are μ and σ^2, respectively.

Problem 1.5 Let $X_1, \cdots, X_n$ be i.i.d. random variables with mean μ and variance σ^2.

- Prove that

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$S_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n)^2,$$

  are unbiased estimators of the mean and variance, that is, $\mathsf{E}[M_n] = \mu$ and $\mathsf{E}[S_n] = \sigma^2$.

- For a distribution of your choice (Gaussian, uniform, etc.) implement these estimators and empirically show on a plot that these estimators indeed converge to the true quantities μ and σ^2 as we increase the sample size n. Based on your plots guess how fast (in terms of powers of n) they converge to the correct quantities.

Problem 1.6 In the mid to late 1980's, in response to the growing AIDS crisis and the emergence of new, highly sensitive tests for the virus, there were a number of calls for widespread public screening for the disease. Similar issues arise in any broad screening problem (e.g., drug testing). The focus at the time was the sensitivity and specificity of the tests at hand. For the tests in question the sensitivity was P(Positive Test | Infected) ≈ 1 and the false positive rate was P(Positive Test | Uninfected) $\approx .00005$ – an unusually low false positive rate. What was generally neglected in the debate, however, was the low prevalence of the disease in the general population: P(Infected) ≈ 0.0001. Since being told you are HIV positive has dramatic ramifications, what clearly matters to you as an individual is the probability that you are uninfected given a positive test result: P(Uninfected | Positive test). Calculate this probability. Would you volunteer for such screening? How does this number change if you are in a "high risk" population – i.e., if P(Infected) is significantly higher?