Boston University
Department of Electrical and Computer Engineering

ENG EC 503 (Ishwar) Learning from Data

Assignment 1

Issued: Thu 19 Jan 2017 **Due:** 5pm Mon 30 Jan 2017 in box outside PHO440
Required reading: Your notes from lectures and additional notes on website.

Problem 1.1 *(Linear Algebra)* 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) The dimension of $\mathbf{v}_1$ is:

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

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

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

(f) If $\text{Proj}_{\mathcal{S}}(\mathbf{v}_3) = a_1\mathbf{v}_1 + a_2\mathbf{v}_2$, 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) Let $B = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}$. Compute: (i) its eigenvalues and (ii) a set of orthonormal eigenvectors.

(h) 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 *(Multivariate Calculus)* 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) 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) Compute the gradient vector $\nabla f(\mathbf{x})$ when $f(\mathbf{x}) = \mathbf{x}^\top A \mathbf{x}$.

(c) 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 *(Two Discrete Random Variables)* Let X and Y be discrete random variables with joint probability mass function (pmf) $p(x,y)$ given by:

$p(x,y)$	$x=-1$	$x=0$	$x=1$
$y=1$	0	$1/8$	0
$y=0$	$1/3$	$1/12$	$1/3$
$y=-1$	0	$1/8$	0

(a) Marginal pmf of X: for $x = -1, 0, 1$, $p(x) =$

(b) Mean/Expectation: $\mu_X = E[X] =$, $\mu_Y = E[Y] =$

(c) Variance: $\sigma_X^2 = \text{var}(X) =$, $\sigma_Y^2 = \text{var}(Y) =$

(d) Correlation: $E[XY] =$ Are X and Y orthogonal? Yes/No, Why?

(e) Covariance: $\text{cov}(X, Y) = E[(X-\mu_X)(Y-\mu_Y)] =$ Are X and Y uncorrelated? Yes/No, Why?

(f) Conditional pmf: $P(X = x|Y = 0)$ for $x = -1, 0, 1$:

(g) Are X and Y independent? Yes/No, Why?

(h) Conditional Mean/Expectation: $E[X|Y = 0] =$

Problem 1.4 *(Bayes Rule)* 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 $P(\text{Positive Test} \mid \text{Infected}) \approx 1$ and the false positive rate was $P(\text{Positive Test} \mid \text{Uninfected}) \approx .00005$ – 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: $P(\text{Infected}) \approx 0.0001$. 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: $P(\text{Uninfected} \mid \text{Positive test})$. 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 $P(\text{Infected})$ is significantly higher?

Problem 1.5 *(Miscellaneous)*

(a) True/False (with reason): If $f_{X,Y}(x, y) = 1$ for all $|x| + |y| \leq 1/\sqrt{2}$ and zero for all other x, y, then X and Y are independent.

(b) True/False (with reason): If $X \sim \mathcal{N}(0, 1)$, Z is independent of X with $P(Z = 1) = 1 - P(Z = -1) = 0.5$, and $Y := XZ$, then X and Y are uncorrelated but not independent.

(c) Let X and Y be IID Bernoulli RVs with $P(X = 0) = P(X = 1) = 0.5$ and $Z := X \oplus Y$ where $\oplus$ denotes modulo-2 addition (XOR). (i) Is Z independent of X? Explain. (ii) Are X and Y conditionally independent given Z? Explain.

(d) Let U, V, W be IID Unif$[-0.5, 0.5]$ RVs. Let $X := W + U$ and $Y := W + V$. (i) Are X and Y independent? Explain. (ii) Are X and Y conditionally independent given W? Explain.

Problem 1.6 (*Working with jointly and conditionally Gaussian random variables*) Let X and Y be jointly Gaussian random variables with means μ_X, μ_Y, variances σ_X^2, σ_Y^2, and correlation coefficient $\rho \in [0, 1]$.

(a) Express $P(aX + bY > 0)$ in terms of the Q-function which is defined by $Q(c) := \frac{1}{\sqrt{2\pi}} \int_c^\infty \exp(-t^2/2)dt$.

(b) If $\mu_X = \mu_Y$, $\sigma_X = \sigma_Y$, and $\rho = 0$, evaluate $P(\{aX + bY > \alpha\} \cap \{bX - aY > \beta\})$ in terms of the Q-function.

(c) If $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$,

  (i) Compute the marginal $f_X(x)$ and conditional $f_{X|Y}(x|y)$ density functions.

  (ii) Express $P(X > 1|Y = y)$ in terms of ρ, y, and the Q-function.

(iii) Express $E[(X - Y)^2 | Y = y]$ in terms of ρ and y.

Problem 1.7 *(Estimating sample complexity)*
We want to estimate the unknown mean μ of a unit-variance scalar Gaussian distribution to an accuracy of ϵ with confidence at least 0.99. What is the minimum number of independent and identically distributed (iid) samples that we will need? Explain how you would go about trying to estimate the minimum number of samples needed.

Problem 1.8 *(Computing orthogonal projections: Approximating* $\sin(t)$ *using polynomials)*
Let

$$x(t) = \begin{cases} \sin(t) & \text{for } -\pi \leq t \leq \pi \\ 0 & \text{otherwise} \end{cases}$$

denote the sine signal restricted to the interval $-\pi \leq t \leq \pi$. For $i = 0, 1, 2, 3, 4, 5$, let

$$b_i(t) = \begin{cases} t^i & \text{for } -\pi \leq t \leq \pi \\ 0 & \text{otherwise} \end{cases}$$

denote the polynomial signals restricted to the interval $-\pi \leq t \leq \pi$. Let $\mathcal{V} = \text{span}\{b_0(t), ..., b_5(t)\}$, that is, $\mathcal{V}$ is the vector space formed by taking all possible linear combinations of the signals $b_0(t), ..., b_5(t)$. We want to find the orthogonal projection of $x(t)$ onto $\mathcal{V}$, that is, we want to find a signal of the form

$$y(t) = a_0 b_0(t) + a_1 b_1(t) + a_2 b_2(t) + a_3 b_3(t) + a_4 b_4(t) + a_5 b_5(t)$$

which is closest to $x(t)$ where closeness is measured with respect to the norm given by

$$||x(t) - y(t)||^2 = \int_{-\infty}^{+\infty} |x(t) - y(t)|^2 dt,$$

that is, the energy of the approximation error.

(a) Using the orthogonality principle of linear algebra explained in the class notes and review session, compute the coefficients $a_0, ..., a_5$ of the orthogonal projection of $x(t)$ onto $\mathcal{V}$. Some of the numerical computations can be done using MATLAB.

(b) Using MATLAB, plot graphs of $\sin(t)$, $y(t)$ and the Madhava-Taylor approximation polynomial $p(t) = t - \frac{t^3}{3!} + \frac{t^5}{5!}$ in the interval $-\pi \leq t \leq \pi$. How well do $y(t)$ and $p(t)$ fare in approximating $\sin(t)$ in the interval $-\pi \leq t \leq \pi$? Specifically, what are the values of $|x(3) - y(3)|$ and $|x(3) - p(3)|$?