Boston University
Department of Electrical and Computer Engineering

ENG EC 503 (Ishwar) Learning from Data

Assignment 1 Solution

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: 3

(b) The length, i.e., norm $\|\mathbf{v}_1\|$, of $\mathbf{v}_1$ is: $\sqrt{1^2 + 1^2 + 0^2} = \sqrt{2}$.

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

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

No, because their inner product is not zero.

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

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) 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 $\mathbf{a} = (a_1, a_2)^\top =$ $(2/3, 2/3)^\top$:

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

(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).

(h) 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.

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

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

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

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,

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

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

(c) 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.

$f(\mathbf{x})$ 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 *(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 $x = -1, 0, 1$, $p(x) =$ $= 1/3$ for all x.

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

We have, $\mu_X = E[X] = -1 \times 1/3 + 0 \times 1/3 + 1 \times 1/3 = 0$. Similarly, $\mu_Y = E[Y] = 0$.

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

We have $\sigma_X^2 = \mathrm{var}(X) = (-1-0)^2 \times 1/3 + (0-0)^2 \times 1/3 + (1-0)^2 \times 1/3 = 2/3$. Similarly, $\mathrm{var}(Y) = 1/4$.

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

$E[XY] = 0$. Yes, X and Y orthogonal because their correlation is equal to zero.

(e) Covariance: $\mathrm{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] =$ Are X and Y uncorrelated? Yes/No, Why?
$\mathrm{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y = 0$. Yes, X and Y uncorrelated because $\mathrm{cov}(X, Y) = 0$.

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

$P(X = -1|Y = 0) = P(X = -1, Y = 0)/P(Y = 0) = (1/3)/(3/4) = 4/9$. Similarly, $P(X = 0|Y = 0) = 1/9$, and $P(X = 1|Y = 0) = 4/9$.

(g) Are X and Y independent? Yes/No, Why? No, because $P(X = -1|Y = 0) \neq P(X = -1)$.

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

$E[X|Y = 0] = -1 \times P(X = -1|Y = 0) + 0 \times P(X = 0|Y = 0) + 1 \times P(X = 1|Y = 0) = 0$.

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?

$\boxed{\text{Answer: } 1/3}$

$$
\begin{aligned}
& P(\text{Uninfected} \mid \text{Positive test}) \\
&= \frac{P(\text{Positive test} \mid \text{Uninfected})P(\text{Uninfected})}{P(\text{Positive test})} \\
&= \frac{P(\text{Positive test} \mid \text{Uninfected})P(\text{Uninfected})}{P(\text{Positive test} \mid \text{Uninfected})P(\text{Uninfected}) + P(\text{Positive test} \mid \text{Infected})P(\text{Infected})} \\
&= \frac{(\text{False Positive Rate})(1 - P(\text{Infected}))}{(\text{False Positive Rate})(1 - P(\text{Infected})) + (\text{Sensitivity})P(\text{Infected})} \\
&= \frac{0.00005(1 - .0001)}{0.00005(1 - .0001) + 1(.0001)} = 0.3333
\end{aligned}
$$

So there is a 1/3 probability that you are actually healthy if you are in a low risk population and your test is positive!

If $P(\text{Infected})$ is significantly higher, the probability you are actually healthy given a positive test result rapidly decreases to essentially zero. For example if $P(\text{Infected}) = 0.001$ (.1% prior probability of infection), $P(\text{Uninfected} \mid \text{Positive test}) = 5\%$ while if $P(\text{Infected}) = 0.01$ (1% prior probability of infection), $P(\text{Uninfected} \mid \text{Positive test}) = .5\%$.

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.

$\boxed{\text{False:}}$ When $X = 1/\sqrt{2}$, the only value that Y can take is 0 but when $X = 0$, Y can take any value from $-1/\sqrt{2}$ to $1/\sqrt{2}$. Hence Y depends on X.

(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.

$\boxed{\text{True:}}$ Being symmetrically distributed, X and Z have zero means. Since X and Z are independent, $E[Y] = E[XZ] = E[X]E[Z] = 0$. Also, $E[XY] = E[X^2 Z] = E[X^2]E[Z] = 0$. Hence X and Y are uncorrelated. They are not, however, independent because $Y = 0$ when $X = 0$ but Y is nonzero if X is nonzero. This shows that Y depends on X.

(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.

(i) Yes: First, for $z = 0, 1$, $P(Z = z) = P(X = 0, Y = z) + P(X = 1, Y = 1 \oplus z) = P(X = 0)P(Y = z) + P(X = 1)P(Y = 1 \oplus z) = 0.5 * 0.5 + 0.5 * 0.5 = 0.5$. Thus, Z is Bernoulli 0.5. Then, for all $z, x \in \{0, 1\}$, $P(Z = z \mid X = x) = P(Y = z \oplus x \mid X = x) = P(Y = z \oplus x) = 0.5 = P(Z = z)$. (ii) No: By interchanging the roles of X and Y in part (i), Z is also independent of Y. Thus $P(X = 0, Y = 1 \mid Z = 0) = 0 \neq 0.25 = P(X = 0 \mid Z = 0) \cdot P(Y = 1 \mid Z = 0)$.

(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.

(i) No: If they were independent, they must be uncorrelated. Since U, V, W all have zero means and are independent, $\text{cov}(X, Y) = E[XY] = E[(W + U)(W + V)] = E[W^2] > 0$ which shows that X and Y are not uncorrelated. (ii) Yes: $f_{XY|W}(x, y|w) = f_{UV|W}(x - w, y - w|w) = f_U(x - w) f_V(y - w) = f_{U|W}(x - w|w) f_{V|W}(y - w|w) = f_{X|W}(x|w) f_{Y|W}(y|w)$. Alternative solution: show that the joint characteristic function factorizes conditioned on W.

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 $Q(c) := \frac{1}{\sqrt{2\pi}} \int_c^\infty \exp(-t^2/2)dt$.

Key facts: (i) Linear transformations of jointly Gaussian random variables are jointly Gaussian. (ii) Jointly Gaussian random variables are completely characterized by their second-order statistics, namely, mean and covariance. (iii) $Q(c) = P(W > c)$ where W is a zero-mean, unit-variance, Gaussian random variable. (iv) If jointly Gaussian random variables are uncorrelated, they are also independent.

Let $Z = aX + bY$ then $Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2)$ where $\mu_Z = a\mu_X + b\mu_Y$ and $\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y$. Then

$$P(aX + bY > 0) = P(Z > 0) = P\left(\frac{(Z - \mu_Z)}{\sigma_Z} > -\frac{\mu_Z}{\sigma_Z}\right) = Q\left(-\frac{\mu_Z}{\sigma_Z}\right) = Q\left(\frac{-(a\mu_X + b\mu_Y)}{\sqrt{a^2\sigma_X^2 + 2ab\rho\sigma_X\sigma_Y + b^2\sigma_Y^2}}\right).$$

(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.

Let $\mu_X = \mu_Y = \mu$ and $\sigma_X = \sigma_Y = \sigma$ and

$$\begin{pmatrix} U \\ V \end{pmatrix} := \begin{pmatrix} a & b \\ b & -a \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Then U, V are jointly Gaussian with means $\mu_U = (a + b)\mu$, $\mu_V = (b - a)\mu$ and variances $\sigma_U^2 = a^2\sigma_X^2 + 2ab\rho\sigma_X\sigma_Y + b^2\sigma_Y^2 = (a^2 + 2ab\rho + b^2)\sigma^2$ and $\sigma_V^2 = b^2\sigma_X^2 - 2ab\rho\sigma_X\sigma_Y + a^2\sigma_Y^2 = (b^2 - 2ab\rho + a^2)\sigma^2$.

Furthermore, $\text{cov}(U, V) = ab\sigma_X^2 + (b^2 - a^2)\rho\sigma_X\sigma_Y - ab\sigma_Y^2 = 0 \Rightarrow U, V$ are uncorrelated jointly Gaussian random variables $\Rightarrow$ they are *independent* Gaussian random variables. Thus,

$$\begin{aligned} P(\{U > \alpha\} \cap \{V > \beta\}) &= P(U > \alpha) \cdot P(V > \beta) \\ &= P\left(\frac{U - \mu_U}{\sigma_U} > \frac{\alpha - \mu_U}{\sigma_U}\right) \cdot P\left(\frac{V - \mu_V}{\sigma_V} > \frac{\beta - \mu_V}{\sigma_V}\right) \\ &= Q\left(\frac{\alpha - (a+b)\mu}{\sigma\sqrt{a^2 + b^2}}\right) \cdot Q\left(\frac{\beta - (b-a)\mu}{\sigma\sqrt{a^2 + b^2}}\right). \end{aligned}$$

(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.

$\boxed{\text{Answer: } X \sim \mathcal{N}(0, 1).}$ Since (X, Y) are jointly Gaussian, X is Gaussian with mean $\mu_X = 0$ and variance $\sigma_X^2 = 1$.

$\boxed{\text{Answer: } f_{X|Y}(x|y) = \mathcal{N}(\rho y, 1 - \rho^2)(x).}$ Since X, Y are jointly Gaussian and have zero means and unit variances, $X|Y = y$ is also Gaussian with mean $E[X|Y] = E[X] + \text{cov}(X, Y)\text{cov}(Y, Y)^{-1}(y - E[Y]) = \rho y$ and variance $\text{cov}(X, X) - \text{cov}(X, Y)\text{cov}(Y, Y)^{-1}\text{cov}(Y, X) = 1 - \rho^2$.

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

$$P(X > 1|Y = y) = P\left(\frac{X - \rho y}{\sqrt{1 - \rho^2}} > \frac{1 - \rho y}{\sqrt{1 - \rho^2}}\middle| Y = y\right) = Q\left(\frac{1 - \rho y}{\sqrt{1 - \rho^2}}\right).$$

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

Since $E[X|Y = y] = \rho y$ and $\text{cov}(X|Y = y) = (1 - \rho^2)$,

$$\begin{aligned} E[(X - Y)^2|Y = y] &= E[((X - \rho y) + (\rho y - y))^2|Y = y] \\ &= E[(X - \rho y)^2|Y = y] + (1 - \rho)^2 y^2 \\ &= \text{cov}(X|Y = y) + (1 - \rho)^2 y^2 \\ &= (1 - \rho^2) + (1 - \rho)^2 y^2. \end{aligned}$$

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.

Solution: Let n be the minimum number of samples needed.

The simple average of n iid samples provides an estimate of the true mean. Let us denote the simple average by Z.

Since the samples are iid and Gaussian, Z is a Gaussian random variable with mean μ and standard deviation $\sigma = 1/\sqrt{n}$. The requirement that the confidence be at least 0.99 means that Z is within an ϵ neighborhood around its mean with probability at least 0.99, i.e.,

$$P(|Z - \mu| \geq \epsilon) < 1 - 0.99 \Leftrightarrow P\left(\frac{|Z - \mu|}{\sigma} \geq \frac{\epsilon}{\sigma}\right) < 0.01 \Leftrightarrow 2Q\left(\frac{\epsilon}{\sigma}\right) < 0.01 \Leftrightarrow \frac{\epsilon}{\sigma} > Q^{-1}(0.005).$$

Thus,

$$\epsilon\sqrt{n} > Q^{-1}(0.005) \Leftrightarrow n > \frac{(Q^{-1}(0.005))^2}{\epsilon^2} \Leftrightarrow n_{\min} = \left\lfloor \frac{(Q^{-1}(0.005))^2}{\epsilon^2} \right\rfloor + 1.$$

Using $Q^{-1}(0.005) \approx 2.5758$, we get

$$n_{\min} = \left\lfloor \frac{6.6349}{\epsilon^2} \right\rfloor + 1.$$

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

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

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

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

denote the polynomial signals restricted to the interval $-\pi \le t \le \pi$. 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 $-\pi \le t \le \pi$. How well do $y(t)$ and $p(t)$ fare in approximating $\sin(t)$ in the interval $-\pi \le t \le \pi$? Specifically, what are the values of $|x(3) - y(3)|$ and $|x(3) - p(3)|$?

Solution:

(a) By the orthogonality principle, for all $i = 0, 1, ..., 5$, $(x(t)-y(t)) \perp b_i(t)$, $\Rightarrow \langle x(t), b_i(t)\rangle = \langle y(t), b_i(t)\rangle$ for all $i = 0, 1, ..., 5$. Since $y(t) = \sum_{j=0}^{5} a_j b_j(t)$, this gives (using the linearity properties of inner products) the following system of six linear equations in the six unknowns $a_0, ..., a_5$:

$$\langle x(t), b_i(t)\rangle \quad = \quad \sum_{j=0}^{5} a_j \langle b_j(t), b_i(t)\rangle, \quad i = 0, ..., 5.$$

All the inner products appearing in the above system of equations are elementary definite integrals which can be easily evaluated in closed analytic form or, alternatively, they can be approximated using MATLAB. Note that each inner product is a number. We can rewrite these system of linear equations using matrix–vector notation as $\mathbf{c} = B\mathbf{a}$ where, for all $i, j = 0, 1, ..., 5$, the entry in row–i of the 6×1 column vector $\mathbf{c}$ is equal to $\langle x(t), b_i(t)\rangle$, the entry in row–$i$ and column–j of the 6×6 matrix B is equal to $\langle b_j(t), b_i(t)\rangle$, and the entry in row–$i$ of the 6×1 column vector $\mathbf{a}$ is equal to a_i. It can be verified that the matrix B is invertible. This shows that the polynomial signals $b_0(t), ..., b_5(t)$ are linearly independent. The coefficients of the orthogonal projection can be obtained as $\mathbf{a} = B^{-1}\mathbf{c}$ using MATLAB. The **unique** solution to the above system of six linear equations in six unknowns is given by $a_0 = a_2 = a_4 = 0$, $a_1 = 0.9879$, $a_3 = -0.1553$, and $a_5 = 0.0056$.

There is another way to deduce that $b_0(t), ..., b_5(t)$ are linearly independent: If $q(t) = \sum_{i=0}^{5} \alpha_i t^i = 0$ for all values of t in the closed interval $[-\pi, \pi]$ then it must be that $\alpha_i = 0$ for all i because otherwise it would contradict the fundamental theorem of Algebra which asserts that $q(t)$ which is polynomial of degree 5 can have no more than 5 roots (real or complex and counting multiple repeated roots).

There is also another way to arrive at the same set of six linear equations in the six unknowns $a_0, ..., a_5$, using elementary (but bit more tedious) calculus of several variables: For each $i = 0, ..., 5$, set

$$\frac{\partial}{\partial a_i} \| \sin(t) - \sum_{i=0}^{5} a_i t^i \|^2 = \frac{\partial}{\partial a_i} \int_{-\infty}^{+\infty} \left| \sin(t) - \sum_{i=0}^{5} a_i t^i \right|^2 dt = 0.$$

This gives a system of six linear equations in the six unknowns $a_0, ..., a_5$ which is exactly the same as the one we obtained before which can be solved in MATLAB.

There is a slightly easier way to compute the coefficients of the orthogonal projection: Since $\sin(t)$ is an odd function of t on the interval $-\pi \le t \le \pi$, the orthogonal projection will live only in the three dimensional subspace of odd symmetric polynomials: $\mathrm{span}\{b_1(t), b_3(t), b_5(t)\}$ (the proof of this assertion is easy). The problem then reduces to the solution of a system of three linear equations in three unknowns along the lines outlined above.

Another completely different (but more roundabout) approach to compute the orthogonal projection is to use the Gram–Schmidt technique on $b_0(t), ..., b_5(t)$ to extract an equivalent *orthonormal* set of signals $\tilde{b}_0(t), ..., \tilde{b}_5(t)$ that spans $\mathcal{V}$. The projection onto $\mathcal{V}$ can then be directly computed by finding the inner products with $\tilde{b}_0(t), ..., \tilde{b}_5(t)$. The polynomial signals that are obtained by orthonormalizing the *elementary* polynomial signals of the form t^i using the Gram–Schmidt technique are the famous Legendre polynomials.

(b) Figure 1 below shows the graphs of $\sin(t)$, $y(t)$ and the Madhava-Taylor polynomial approximation $p(t) = t - \frac{t^3}{3!} + \frac{t^5}{5!}$ over the interval $-\pi \le t \le \pi$. The approximation $y(t)$ is so accurate that the graphs of $y(t)$ and $\sin(t)$ are almost identical. The Madhava-Taylor polynomial is an excellent approximation to $\sin(t)$ for t near 0. But for $|t| > 2$, the Madhava-Taylor polynomial deviates significantly from $\sin(t)$. For example the Madhava-Taylor polynomial approximation error at $t = 3$ is about 0.3839 while the approximation error of $y(t)$ is only about 0.001416, which is about 271 times smaller than the approximation error of the Madhava-Taylor polynomial!

Detailed derivation: For i, j, both odd positive integers,

$$B_{ij} := \langle t^i, t^j \rangle = \int_{-\pi}^{\pi} t^{i+j} dt = \left. \frac{t^{i+j+1}}{(i+j+1)} \right]_{-\pi}^{\pi} = \frac{2\pi^{i+j+1}}{(i+j+1)}.$$

$$B_{11} = \frac{2\pi^3}{3},$$

$$B_{12} = B_{21} = \frac{2\pi^5}{5},$$

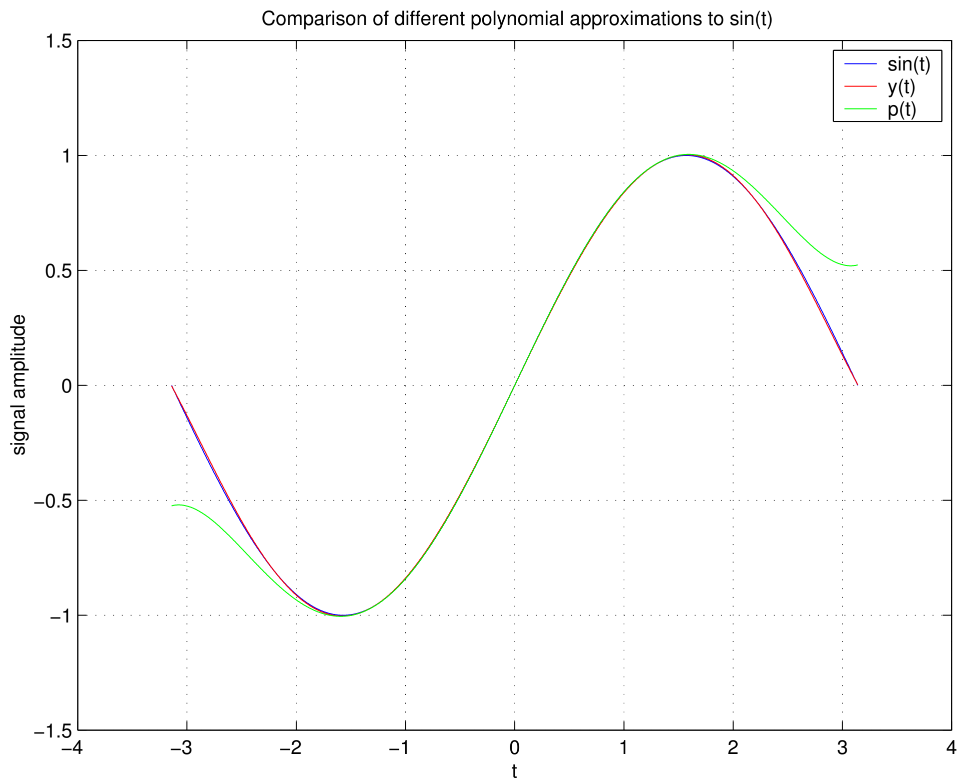

Figure 1: Graphs of $\sin(t)$, $y(t)$, and $p(t)$ in the interval $-\pi \le t \le \pi$.

$$B_{13} = B_{22} = B_{31} = \frac{2\pi^7}{7},$$

$$B_{23} = B_{32} = \frac{2\pi^9}{9},$$

$$B_{33} = \frac{2\pi^{11}}{11}.$$

$$B = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{13} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix}$$

For j an odd positive intger,

$$\begin{aligned} c_j &= \langle t^j, \sin(t) \rangle \\ &= 2\int_0^\pi t^j \sin(t)dt. \\ &= -2t^j \cos(t)\Big]_0^\pi + 2\int_0^\pi jt^{j-1}\cos(t)dt \\ &= 2\pi^j + j\left[\left(2t^{j-1}\sin(t)\right)_0^\pi - 2\int_0^\pi (j-1)t^{j-2}\sin(t)dt\right] \\ &= 2\pi^j + j\left[-(j-1)2\int_0^\pi t^{j-2}\sin(t)dt\right] \\ &= 2\pi^j - j(j-1)c_{j-2}. \end{aligned}$$

For $j = 1$,

$$c_1 = 2\int_0^{\pi} t\sin(t)dt = 2\left[(-t\cos(t))_0^{\pi} + \int_0^{\pi}\cos(t)dt\right] = 2[\pi + 0] = 2\pi.$$

Thus, $\boxed{c_1 = 2\pi,}$

$c_3 = 2\pi^3 - 6c_1 = \boxed{2\pi^3 - 12\pi}$, and

$c_5 = 2\pi^5 - 20c_3 = 2\pi^5 - 20(2\pi^3 - 12\pi) = \boxed{2\pi^5 - 40\pi^3 + 240\pi}$.

$$\mathbf{c} = (c_1, c_3, c_5)^{\top}$$

$$\mathbf{a} = B^{-1}\mathbf{c} = (0.9879, -0.1553, 0.0056)^{\top}.$$