Problem 1 *10 points*

For each of the following parts, indicate whether the statement is always true or it can be false by clearly writing "True" or "False." Briefly explain the reasoning behind your answer for partial credit (in case your choice is wrong). Diagrams are welcome.

(a) If under hypothesis H_0, Y is Gaussian $(0, 1)$ and, under hypothesis H_1, Y is Gaussian $(1, 1)$, then we will always decide H_0 when $y < 1/2$ under the MAP rule.

False. This would be true for the ML detector, but the MAP detector will shift with the prior probabilities $\mathbb{P}[H_0], \mathbb{P}[H_1]$.

(b) Let $Y = X + Z$ where X and Z are independent Gaussian $(0, 1)$ random variables. Then, $\hat{x}_{\text{MMSE}}(y) = y/2$.

True. By computation, $\mathsf{Cov}[X, Y] = 1$, $\mathsf{Var}[Y] = 1 + 1 = 2$, and $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. The MMSE equation for Gaussians reduces to the statement.

(c) For a Markov chain, X_0 (the state at time 0) is independent of X_2 (the state at time 2).

False. The whole purpose of Markov chains is to model correlations between states at different times.

(d) Suppose we have a sequence X_i of independent, identically distributed uniform random variables, each uniform on the interval [-3,3]. Define the normalized sums $S_n = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i}{\sqrt{3}}$. Then, according to the Central Limit Theorem, as $n \to \infty$, $P(S_n > 1) \to 1 - \Phi(1) = \frac{1}{\sqrt{2\pi}} \int_1^\infty e^{-\frac{x^2}{2}} \, dx$.

False. Wrong scaling with respect to n for the central limit theorem.

(e) In a machine learning problem, the training error is always less than or equal to the test error.

False. This is only true for some classifiers that overtrain, such as the nearest-neighbor classifier.

Problem 2 *10 points*

For each of the following parts, calculate the quantity of interest. Show your steps for partial credit.

(a) (2 pts) Suppose we have a detection problem such that the observation Y, has conditional density $P_{Y|H_0}(y)$ that is uniformly distributed on $[-1, 1]$. Suppose we have a decision rule

$$D(y) = \begin{cases} \text{Say } H_1 & |y| < 2/3 \\ \text{Say } H_0 & \text{elsewhere.} \end{cases}$$

What is the probability of false alarm?

$P_{FA} = 2/3$.

(b) (4 pts) Two jointly continuous random variables X and Y have $E[X] = 2$, $E[Y] = 1$, $\mathsf{Var}[X] = \mathsf{Var}[Y] = 1$ and $\mathsf{Cov}(X, Y) = -0.5$. Compute $\widehat{x}_{LLSE}(y)$, the linear least squares estimator of X given Y. Let $e = X - \widehat{x}_{LLSE}(Y)$. Compute $\mathbb{E}[e^2]$, the mean square error of the estimate.

$\mathbb{E}[X|Y] = 2 - \frac{0.5}{1}(Y - 1)$. $\mathbb{E}[e] = 0$ (unbiased). $\mathbb{E}[e^2] = Var(X) - \frac{Cov(X,Y)^2}{Var(Y)} = 1 - 1/4 = 3/4$.

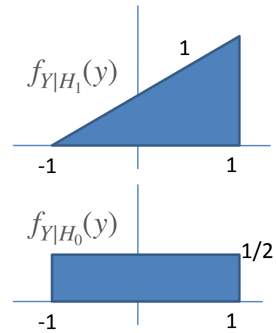
(c) (4 pts) Let $X_1, X_2, \ldots, X_{48}$ be independent uniform random variables in $[-2, 2]$. and let $Y = X_1 + X_2 + \cdots + X_{48}$. Let Φ denote the CDF of a standard normal random variable. Compute the variance of Y, and estimate the probability that $Y \leq 20$ in terms of Φ using the Central Limit Theorem.

$\mathsf{Var}[X_i] = \frac{4^2}{12} = \frac{4}{3}$. $\mathsf{Var}[Y] = 48\mathsf{Var}[X_i] = 64$. Standard deviation of Y is 8. $\mathbb{E}[Y] = 0$. So, by CLT $\mathbb{P}[Y \leq 20] = \Phi(2.5)$.

Problem 3 *16 points*

A random variable Y is observed, and has two possible distributions (hypothesis H_1 or hypothesis H_0), summarized as

- When H_0 is true, Y is uniformly distributed on (-1, +1).
- When H_1 is true, Y has pdf $f_{Y|H_1}(y) = \begin{cases} \frac{1+y}{2}, & -1 < y < +1 \\ 0 & \text{elsewhere} \end{cases}$



The situation is illustrated in the adjacent figure. You are to design detectors based on observing Y to select hypothesis H_1 or h_0. **For this problem, answers in terms of explicit numbers are required for full credit.**

(a) Suppose that the maximum-likelihood (ML) decision rule is being used. What value(s) of Y result in a decision in favor of H_1?

Clearly, H_1 has higher probability density for $0 \leq x \leq 1$.

(b) Find the false alarm probability P_{FA} and the probability of missed detection P_{MD} of the maximum-likelihood decision rule.

$P_{FA} = 0.5$, the area of the rectangle from 0 to 1. $P_{MD} = 0.25$, the area of the triangle from -1 to 1.

(c) Now suppose that $\mathbb{P}[H_0] = \frac{1}{4}, \mathbb{P}[H_1] = \frac{3}{4}$. For the Maximum a Posteriori (MAP) decision rule, what value(s) of Y result in a decision in favor of H_1 ?

Threshold is 1/3, so we want

$$y + 1 = 1/3 \implies y = -2/3$$

as the boundary between the two regions. Thus, H_1 is selected for $-2/3 \leq y \leq 1$.

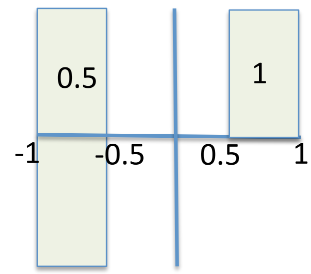
(d) Compute the average error probability P_e^{MAP} for the MAP decision rule when $\mathbb{P}[H_0] = \frac{1}{4}, \mathbb{P}[H_1] = \frac{3}{4}$.

$P_{FA} = 5/6$, the area under the rectangle for $-2/3 < y < 1$. $P_{MD} = 1/36$, the area under the triangle for $-1 < y < -2/3$. $P_e^{MAP} = 3/4 \times 1/36 + 1/4 \times 5/6 = 11/48$

Problem 4 *16 points*

Let X, Y be joint random variables, with *non-uniform* joint probability density in the region delineated in the figure below, given by:

$$f_{X,Y}(x,y) = \begin{cases} 1 & 0.5 \le x \le 1, 0 \le y \le 1 \\ 0.5 & -1 \le x \le -0.5, -1 \le y \le 1 \\ 0 & else \end{cases}$$



(a) Compute $\mathbb{E}[X]$ as a number (must do the integral or use any other way)

$$\mathbb{E}[X] = \int_{0.5}^{1}\int_{0}^{1} x\,dy\,dx + \int_{-1}^{-0.5}\int_{-1}^{1} x\,dy\,dx = 0$$

Also obvious by symmetry.

(b) Note that the marginal density $f_Y(y)$ is given by

$$f_Y(y) = \int_{-1}^{1} f_{X,Y}(x,y)dx = \begin{cases} 0.75 & 0 \le y \le 1 \\ 0.25 & -1 \le y < 0 \\ 0 & else \end{cases}$$

Compute $\mathbb{E}[Y]$ as a number (must do the integral or use any other way)

$$\mathbb{E}[Y] = \int_{-1}^{1} y f_y(y)dy = \int_{-1}^{0} y/4dy + \int_{0}^{1} 3y/4dy = -1/8 + 3/8 = 1/4$$

(c) Compute the MMSE estimate of X given the observation that $Y = -0.5$. You may want to start by computing the conditional PDF $f_{X|Y}(x|-0.5)$.

When $Y = -0.5$, X is restricted to a uniform random variable in $[-1, -0.5]$, so $\mathbb{E}[X|Y = -0.5] = -0.75$.

(d) Note that $\mathbb{E}[Y^2] = 1/3$ and $\mathbb{E}[XY] = 3/16$. Compute the best linear estimator $\hat{X}_{LLSE}(Y)$ for all values of Y.

$$\hat{x}_{LLSE}(Y) = \mathbb{E}[X] + \frac{\mathsf{Cov}(X,Y)}{\mathsf{Var}(Y)}(Y - \mathbb{E}[Y]) = \frac{\frac{3}{16}}{\frac{13}{48}}(Y - \frac{1}{4}) = \frac{9}{13}(Y - \frac{1}{4})$$

Problem 5 *16 points*

A biomedical engineer is testing whether their new device improves patient outcomes after knee surgery. They know that the typical patient population with this condition has a symptom severity rating of 7.2 with a standard deviation of 1. As a preliminary study, they examine 100 patients who have received the device and find a reported severity rating of 6.5 with a sample standard deviation of 1.5. Here are some useful pre-calculated values of the Φ function:

- $\Phi(-1) = 0.16, \Phi(-1.3) = 0.1, \Phi(-1.6) = 0.05,\ \Phi(-2) = 0.025, \Phi(-3) = 0.0013$
- $\Phi(-4) = 3.2 \times 10^{-5}, \Phi(-5) = 2.9 \times 10^{-7}, \Phi(-6) = 9.7 \times 10^{-10}, \Phi(-7) = 1.3 \times 10^{-12}$

(a) Let $X_i, i = 1, \dots, 100$ denote the severity rating of patient i, and let $M_{100} = \frac{1}{100}\sum_{i=1}^{100} X_i$ denote the average response of the 100 patients. If we assume the variance of each X_i is 2.25, what is the standard deviation of M_{100}?

The variance of M_{100} is $\frac{2.25}{100}$. Hence the standard deviation is

$$\sqrt{\frac{2.25}{100}} = 0.15.$$

(b) Construct the 90% confidence interval for the true average severity rating of patients with the new device, centered at the sample mean M_{100}. Note that there are enough patients to assume M_{100} is approximately Gaussian, with standard deviation as computed in part (a).

The standard deviation of M_{100} is 0.15. Since $\Phi(-1.6) = 0.05$, the 90% confidence interval around the sample mean is $[6.5 - 1.6 \times 0.15, 6.5 + 1.6 \times 0.15] = [6.26, 6.74]$. Hence, the true mean 7.2 is outside the 90% confidence interval around the sample mean.

(c) Assume you know the true mean and variance of symptom severity of normal patients. Explain the null hypothesis, and the test you would apply to the data to identify if the device improves patient outcomes (selecting from one of the tests we covered in lecture).

The null hypothesis is that the new device does not improve patient outcomes as measured by symptom severity rating. Since the population standard deviation is known, we can apply a one sample Z-test for difference in means.

(d) Based on the observed test data, should the null hypothesis be rejected with significance level 0.05? Explain all steps in your calculation.

First, compute the Z-statistic assuming the known mean under the null hypothesis (7.2) and the standard deviation of 1. This is

$$Z = \frac{\sqrt{n}(M_n - \mu)}{\sigma} = \frac{\sqrt{100}(6.5 - 7.2)}{1} = 10 \cdot -0.7 = -7.$$

We can now use one of the pre-computed $\Phi(z)$ values above to compute the p-value:

$$\mathrm{p-value} = 2\Phi(-|Z|) = 2\Phi(-7) = 2.6 \times 10^{-12}$$

This is much smaller than the significance level of 0.05, so we reject the null hypothesis and conclude that the device has an effect.

Problem 6 *16 points*

In this problem, you will work through the process of constructing and evaluating an LDA binary classifier by hand. (All of the values below were carefully chosen to make the calculations easy since MATLAB and other computing software is not allowed.) You have been given the following 1-dimensional training and test datasets:

$$\mathbf{X}_{\text{train}} = \begin{bmatrix} +3 \\ +2 \\ +1 \\ 0 \\ -1 \\ -2 \end{bmatrix} \quad \underline{Y}_{\text{train}} = \begin{bmatrix} +1 \\ -1 \\ -1 \\ -1 \\ +1 \\ +1 \end{bmatrix}$$

$$\mathbf{X}_{\text{test}} = \begin{bmatrix} 2.5 \\ 0.5 \\ -0.5 \\ -1.5 \end{bmatrix} \quad \underline{Y}_{\text{test}} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix}$$

(a) Compute the sample means of the training data $\hat{\mu}_+$ and $\hat{\mu}_-$ as well as the sample covariance matrix $\hat{\mathbf{\Sigma}}$, which in this 1-dimensional setting is just a sample variance (and could be denoted by $\hat{\sigma}^2$ instead if you wish). Show your work for full credit.

$$\mu_+ = (3-1-2)/3 = 0; \quad \mu_- = (2+1+0)/3 = 1$$
$$\Sigma_+ = \left((3-0)^2 + (-1-0)^2 + (-2-0)^2\right)/(3-1) = 7$$
$$\Sigma_- = \left((2-1)^2 + (1-1)^2 + (0-1)^2\right)/(3-1) = 1$$
$$\Sigma = \frac{1}{3+3-2}(2\Sigma_+ + 2\Sigma_-) = \frac{1}{4}(14+2) = 4.$$

(b) Suppose you design the nearest neighbor classifier. Compute the sample probability of error of the nearest neighbor classifier on the test data.

This is a bit odd, because there are ties for the nearest neighbor. This was a bad choice of points, so here are acceptable answers:

$P_e = 0,$ when 2.5 is associated with neighbor 3, and -0.5 is associated with neighbor -1

$P_e = 0.5,$ when 2.5 is associated with neighbor 2, and -0.5 is associated with neighbor 0

$P_e = 1/4,$ when 2.5 is associated with 2, but -0.5 is associated with -1.

$P_e = 1/4,$ when 2.5 is associated with 3, but -0.5 is associated with 0.

Bottom line: all of the above answers are appropriate.

(c) Suppose you design the closest average classifier using the averages computed from part (a). Compute the performance of this classifier on the training and test data sets.

On the training data, the closest average selects decisions

$$D(y) = \begin{cases} -1 & y \geq 1/2 \\ 1 & y < 1/2 \end{cases}$$

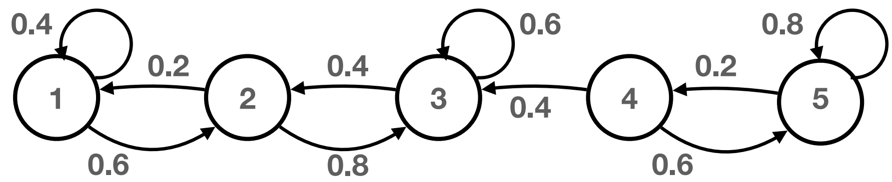
Then, the probability of error error on the training set is $\frac{1}{3}$, as $y = 3$ and $y = 0$ are misclassified.

On the test set, the probability of error is a bit ambiguous due to a tie at $y = 1/2$. We can accept answers of $1/4$ or $1/2$, depending on how we break the tie.

(d) Design the Linear Discriminant Analysis classifier (LDA) for this problem. Note that you don't need to know the actual value of the sample variance to determine the classifier. Evaluate the performance of this classifier on the training and test data.

With little thinking, LDA is a classifier assuming Gaussian models for the data with different means μ_+, μ_- and a common covariance Σ. This classifier is exactly the same as the closest average, which has training error $\frac{1}{3}$, and test error either $1/4$ or $1/2$, depending on how one breaks the tie at $y = 0.5$.

Problem 7 *16 points*



Consider a Markov chain with the transitions probability graph as drawn above.

(a) Write the state transition matrix $[P_{ij}]$, where $P_{ij} = P(x(t+1) = j | x(t) = i)$.

$$P = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 & 0 \\ 0.2 & 0 & 0.8 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.4 & 0 & 0.6 \\ 0 & 0 & 0 & 0.2 & 0.8 \end{pmatrix}$$

(b) Find the transient states and the recurrent states.

States 1, 2, 3 are recurrent, states 4 and 5 are transient because there is a path to a recurrent state, with no return possible for those two states.

(c) Is the Markov chain irreducible? Is the Markov chain aperiodic? Explain.

Not irreducible, because there are transient states. It is aperiodic, because the recurrent communicating class has period 1 (two self-loops)

(d) Does a unique limiting state probability vector exist? If so, argue why and solve for it. If not, argue why.

Throw out the transient states, this looks like a three state chain. Flow balance says

$$0.2pi_2 = 0.6\pi_1; \quad 0.4\pi_3 = 0.8\pi_2$$

which simplifies to $pi_2 = 3\pi_1; \pi_3 = 2\pi_2 = 6\pi_1$. Conservation of probability:

$$\pi_1 + \pi_2 + \pi_3 = 1 = (1 + 3 + 6)\pi_1$$

This means $\pi_1 = 1/10, \pi_2 = 3/10, \pi_3 = 6/10$.