CIS3333: Mathematics of Machine Learning Fall 2025

Homework 3

Due: September 26, 2025, 11:59 PM ET

Submission Instructions: Submit a single PDF (clear scans or photos compiled) to Gradescope and assign pages for each problem. Show key steps and justify answers.

Collaboration & AI Policy: You may discuss approaches with classmates, but write up your own solutions and list collaborators. If you use computational tools (including LLMs) for checking, cite them and ensure the reasoning is your own.

Problem 1: Bayesian Updating (10 points)

You are a data scientist at a music streaming service, and you're testing a new song to see if it has viral potential. You model the probability that a random user will "like" the song, μ, with a Beta distribution. You have some information you've gathered from previous songs, so you start with a prior, $p(\mu) = \text{Beta}(\mu|\alpha, \beta)$. Now, you start getting data, one user at a time, where $x_i = 1$ if the user likes the song and $x_i = 0$ if they don't.

a. (1 point) First, write down the probability mass function (PMF) for the likelihood of a single observation, $p(x_i|\mu)$. What is the name of this distribution?

b. (4 points) The first user listens to the song, providing feedback x_1. Starting with your initial prior, show that the posterior distribution $p(\mu|x_1)$ is also a Beta distribution. Derive its parameters, α_1 and β_1.

c. (3 points) Now, a second user provides feedback x_2. Using the posterior from the first user as your new prior, show that the posterior after observing x_2, $p(\mu|x_1, x_2)$, is again a Beta distribution. What are its parameters, α_2 and β_2?

d. (2 points) After N users have given their feedback, you have a sequence of likes and dislikes, $x_1, \ldots, x_N$. Based on the pattern you observed, what is the final posterior distribution $p(\mu|x_1, \ldots, x_N)$? Express the final posterior parameters (α_N and β_N) in terms of the initial prior parameters (α, β) and the observed data $x_1, \ldots, x_N$.

Problem 2: Gaussian Conditional Distribution (10 points)

The probability density function for a zero-mean bivariate Gaussian with unit variances ($\mu_1 = \mu_2 = 0, \sigma_1^2 = \sigma_2^2 = 1$) and correlation ρ is given by:

$$p(x_1, x_2) \propto \exp\left(-\frac{1}{2(1-\rho^2)}\left[x_1^2 + x_2^2 - 2\rho x_1 x_2\right]\right)$$

a. (3 points) Show that the marginal distribution of x_2 is $\mathcal{N}(0, 1)$.

b. (5 points) Use the above and the formula for the joint distribution to show that the conditional distribution of x_1 given x_2 is a Gaussian with mean ρx_2 and variance $1 - \rho^2$, i.e., $x_1 | x_2 \sim \mathcal{N}(\rho x_2, 1 - \rho^2)$.

c. (1 point) For random variables in general, zero correlation does not imply independence. Show that for the special case of a bivariate Gaussian, zero correlation ($\rho = 0$) *does* imply independence.

d. (1 point) Using your result for the conditional variance $\sigma^2_{1|2}$, show that conditioning reduces variance, i.e., $\sigma^2_{1|2} \leq 1$. When does equality hold?

Problem 3: Concentration for Exponential Variables (10 points)

The Chernoff method is a general recipe for deriving exponential tail bounds. In this problem, you will apply it to a single Exponential random variable. An exponential random variable with rate λ has the probability density function $p(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.

a. (2 points) Show that the mean of this distribution is $\mathbb{E}[X] = 1/\lambda$. (Hint: You may need to use integration by parts).

b. (3 points) Show that the Moment Generating Function (MGF) is $M_X(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$.

c. (5 points) For simplicity, let's now consider the case where $\lambda = 1$, so the mean is 1. Using the Chernoff bound, prove the following concentration inequality for any $\epsilon > 0$:

$$\mathbb{P}\left(X \geq 1 + \epsilon\right) \leq (1 + \epsilon) e^{-\epsilon}$$

(Hint: Apply the Chernoff bound and find the optimal $t \in (0, 1)$ to make the bound as tight as possible).