

Homework 1

Math for AI

Instructions: Please solve the following problems showing all your work. State any assumptions you make.

1. Data Analysis

You are analyzing the performance of a new classification model. Let X be the random variable representing the model's confidence score (discretized into Low, Medium, High), and let Y be the variable representing whether the prediction was Correct ($Y = 1$) or Incorrect ($Y = 0$). The joint probability distribution $P(X, Y)$ is given below:

Prediction (Y)	Confidence (X)		
	Low (x_1)	Medium (x_2)	High (x_3)
Incorrect ($Y = 0$)	0.15	0.10	0.05
Correct ($Y = 1$)	0.05	0.15	0.50

- (a) Calculate the marginal distribution of the confidence scores, $P(X)$.

Solution: We sum across the rows for each column of X :

$$P(X = \text{Low}) = 0.15 + 0.05 = 0.20$$

$$P(X = \text{Medium}) = 0.10 + 0.15 = 0.25$$

$$P(X = \text{High}) = 0.05 + 0.50 = 0.55$$

- (b) What is the overall accuracy of the model, $P(Y = 1)$?

Solution: We sum across the row for $Y = 1$:

$$P(Y = 1) = 0.05 + 0.15 + 0.50 = 0.70$$

The overall accuracy is 70%.

- (c) Calculate the conditional probability that a prediction is correct given that the confidence is High, $P(Y = 1|X = \text{High})$.

Solution: Using the definition of conditional probability:

$$P(Y = 1|X = \text{High}) = \frac{P(Y = 1, X = \text{High})}{P(X = \text{High})} = \frac{0.50}{0.55} = \frac{10}{11} \approx 0.9091$$

- (d) Are the random variables X and Y independent? Justify your answer mathematically.

Solution: For X and Y to be independent, $P(X, Y) = P(X)P(Y)$ must hold for all combinations. Let's check $X = \text{Low}$ and $Y = 1$:

$$P(X = \text{Low}, Y = 1) = 0.05$$

$$P(X = \text{Low})P(Y = 1) = (0.20)(0.70) = 0.14$$

Since $0.05 \neq 0.14$, X and Y are **not independent**.

2. System Reliability

A cloud computing request must pass through three independent layers of security checks to be processed: a Firewall (F), an Authentication Service (A), and an Authorization Policy (Z).

- The probability the Firewall passes the request is 0.98
- The probability the Authentication Service passes the request given that the Firewall has passed is 0.95.
- The probability the Authorization Policy passes the request given that both the Firewall and the Authentication Service have passed is 0.99.

- (a) What is the probability that a legitimate request is successfully processed (passes all three checks)?

Solution: We want the joint probability $P(F \cap A \cap Z)$. Using the chain rule of probability:

$$\begin{aligned}P(F \cap A \cap Z) &= P(F) \times P(A|F) \times P(Z|F \cap A) \\P(F \cap A \cap Z) &= 0.98 \times 0.95 \times 0.99 = 0.92169\end{aligned}$$

- (b) What is the probability that a request is rejected specifically by the Authentication Service?

Solution: For a request to be rejected by the Authentication Service, it must first pass the Firewall (F) and then fail the Authentication check (A^c). We are looking for the joint probability $P(F \cap A^c)$. Using the definition of conditional probability:

$$P(F \cap A^c) = P(F) \times P(A^c|F)$$

Since $P(A|F) = 0.95$, the complement is $P(A^c|F) = 1 - 0.95 = 0.05$.

$$P(F \cap A^c) = 0.98 \times 0.05 = 0.049$$

The probability that a request is rejected specifically by the Authentication Service is 0.049 (or 4.9%).

3. Spam Detection

Let S denote the event that an email is spam ($S = 1$) or not ($S = 0$). Based on prior data, the probability that an incoming email is spam is $P(S = 1) = 0.4$. You have a spam filter that looks for the keyword "Congratulations". Let $K = 1$ indicate the presence of the keyword.

- If an email is spam, the probability it contains the keyword is $P(K = 1|S = 1) = 0.25$.
- If an email is not spam, the probability it contains the keyword is $P(K = 1|S = 0) = 0.05$.

If an incoming email contains the word "Congratulations", what is the posterior probability that it is actually spam?

Solution: We need to find $P(S = 1|K = 1)$. By Bayes' Theorem:

$$P(S = 1|K = 1) = \frac{P(K = 1|S = 1)P(S = 1)}{P(K = 1)}$$

Using the Law of Total Probability to find the denominator $P(K = 1)$:

$$P(K = 1) = P(K = 1|S = 1)P(S = 1) + P(K = 1|S = 0)P(S = 0)$$

$$P(K = 1) = (0.25)(0.4) + (0.05)(1 - 0.4) = 0.10 + 0.03 = 0.13$$

Now, apply Bayes' Theorem:

$$P(S = 1|K = 1) = \frac{0.10}{0.13} = \frac{10}{13} \approx 0.7692$$

4. Attracting events

Prove that for two events in a sample space, if $P(A|B) > P(A)$ then $P(B|A) > P(B)$.

Solution: By definition, $P(A|B) = \frac{P(A \cap B)}{P(B)}$. We are given that $P(A|B) > P(A)$. Therefore:

$$\frac{P(A \cap B)}{P(B)} > P(A) \implies P(A \cap B) > P(A)P(B)$$

Now, look at $P(B|A)$:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Since we know $P(A \cap B) > P(A)P(B)$, we can substitute this inequality:

$$P(B|A) > \frac{P(A)P(B)}{P(A)}$$

$$P(B|A) > P(B)$$

Which completes the proof.

5. Engines

Customers who purchase a certain make of car can order an engine in any of three sizes. Of all cars sold, 45% have the smallest engine, 35% have the medium sized one, and 20% have the largest. Of cars with the smallest engine, 10% fail an emissions test within the warranty period, while 12% of those with medium size and 15% of those with largest engine fail. A friend of yours tells you that her car, which is of this make, failed an emissions test within the warranty period. Compute the probabilities that your friend's car has small, medium or large engine.

Solution: Let S, M, L represent the engine sizes, and F represent the event of failing the emissions test. Given probabilities: $P(S) = 0.45, P(M) = 0.35, P(L) = 0.20$ $P(F|S) = 0.10, P(F|M) = 0.12, P(F|L) = 0.15$

First, find the total probability of failure $P(F)$:

$$P(F) = P(F|S)P(S) + P(F|M)P(M) + P(F|L)P(L)$$

$$P(F) = (0.10)(0.45) + (0.12)(0.35) + (0.15)(0.20) = 0.045 + 0.042 + 0.030 = 0.117$$

Now, use Bayes' Theorem to find the posteriors:

$$P(S|F) = \frac{P(F|S)P(S)}{P(F)} = \frac{0.045}{0.117} = \frac{45}{117} = \frac{5}{13} \approx 0.3846$$

$$P(M|F) = \frac{P(F|M)P(M)}{P(F)} = \frac{0.042}{0.117} = \frac{42}{117} = \frac{14}{39} \approx 0.3590$$

$$P(L|F) = \frac{P(F|L)P(L)}{P(F)} = \frac{0.030}{0.117} = \frac{30}{117} = \frac{10}{39} \approx 0.2564$$

6. Server Pings

A monitoring node sends "pings" to a remote server. Each ping has a probability $p = 0.8$ of receiving a successful echo reply, independent of other pings. (Note: using p instead of μ for standard probability notation).

- (a) Let X be a random variable for a single ping ($X = 1$ for success, $X = 0$ for failure). What is the variance of X ?

Solution: X follows a Bernoulli distribution with $p = 0.8$.

$$\text{Var}(X) = p(1 - p) = 0.8(0.2) = 0.16$$

- (b) You send a batch of $N = 10$ pings. Let Y be the number of successful replies. What kind of distribution does Y follow?

Solution: Since Y is the sum of N independent Bernoulli trials, Y follows a **Binomial distribution**, $Y \sim \text{Binomial}(N = 10, p = 0.8)$.

- (c) Calculate the probability of receiving exactly 8 successful replies.

Solution: Using the Binomial probability mass function:

$$P(Y = 8) = \binom{10}{8} (0.8)^8 (0.2)^2 = 45 \times 0.16777216 \times 0.04 \approx 0.3020$$

- (d) What is the expected number of successful replies in the batch?

Solution: For a Binomial distribution, the expected value is:

$$E[Y] = Np = 10 \times 0.8 = 8$$

7. API Traffic

Requests arrive at a microservice API at an average rate of $\lambda = 4$ requests per second. Assume the arrival of requests follows a Poisson process.

- (a) Write the expression for the probability of receiving k requests in a given second.

Solution: $P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} = \frac{e^{-4} 4^k}{k!}$

- (b) What is the probability that the service receives exactly 0 requests in a specific second?

Solution: $P(X = 0) = \frac{e^{-4} 4^0}{0!} = e^{-4} \approx 0.0183$

- (c) What is the probability that the service receives 2 or more requests in a specific second?

Solution: $P(X \geq 2) = 1 - P(X < 2) = 1 - (P(X = 0) + P(X = 1))$

$$P(X = 1) = \frac{e^{-4} 4^1}{1!} = 4e^{-4}$$

$$P(X \geq 2) = 1 - (e^{-4} + 4e^{-4}) = 1 - 5e^{-4} \approx 1 - 5(0.0183) = 1 - 0.0915 = 0.9085$$

8. LLM Token generation

A simplified Large Language Model has a vocabulary of only 3 tokens: {"cat", "dog", "fish"}. For a specific prompt, the model outputs a categorical distribution with probabilities:

$$\mu_{\text{cat}} = 0.5, \quad \mu_{\text{dog}} = 0.3, \quad \mu_{\text{fish}} = 0.2$$

- (a) If we sample $N = 50$ tokens independently, what is the expected number of times the token "dog" will appear? What is the variance for this count?

Solution: Let X_{dog} be the random variable representing the number of times the token "dog" appears in $N = 50$ independent trials. Since each trial is independent and has a constant probability $p = \mu_{\text{dog}} = 0.3$, X_{dog} follows a Binomial distribution: $X_{\text{dog}} \sim \text{Binomial}(N = 50, p = 0.3)$.

The expected value is calculated as:

$$E[X_{\text{dog}}] = N \cdot p = 50 \cdot 0.3 = 15$$

The variance is calculated as:

$$\text{Var}(X_{\text{dog}}) = N \cdot p \cdot (1 - p) = 50 \cdot 0.3 \cdot (1 - 0.3) = 50 \cdot 0.3 \cdot 0.7 = 10.5$$

- (b) We sample $N = 5$ tokens independently from this distribution. What is the probability of generating exactly: 3 "cats", 1 "dog", and 1 "fish"?

Solution: This is a multinomial distribution problem:

$$P(3 \text{ cats}, 1 \text{ dog}, 1 \text{ fish}) = \frac{5!}{3!1!1!} (0.5)^3 (0.3)^1 (0.2)^1$$

$$P = 20 \times (0.125) \times (0.3) \times (0.2) = 20 \times 0.0075 = 0.15$$

9. Expectation and Variance of a Custom Variable

A GPU cluster job has three possible states after execution:

- State A: Success (Probability 0.6). Reward $R = 10$.
- State B: Soft Failure (Probability 0.3). Reward $R = 1$ (partial results).
- State C: Hard Failure (Probability 0.1). Reward $R = -5$ (wasted compute time).

- (a) Calculate the Expected Reward $E[R]$.

Solution:

$$E[R] = \sum r_i P(R = r_i) = (10)(0.6) + (1)(0.3) + (-5)(0.1)$$

$$E[R] = 6 + 0.3 - 0.5 = 5.8$$

- (b) Calculate the Variance of the Reward $\text{Var}(R)$.

Solution: First calculate $E[R^2]$:

$$E[R^2] = \sum r_i^2 P(R = r_i) = (100)(0.6) + (1^2)(0.3) + ((-5)^2)(0.1)$$

$$E[R^2] = 60 + 0.3 + 2.5 = 62.8$$

Then, $\text{Var}(R) = E[R^2] - (E[R])^2$:

$$\text{Var}(R) = 62.8 - (5.8)^2 = 62.8 - 33.64 = 29.16$$

10. **Maximum Likelihood Estimation**

Derive the formula for the MLE of the parameter λ of the Poisson distribution.

Solution: Let x_1, x_2, \dots, x_n be a random sample of size n from a Poisson distribution with parameter λ . The likelihood function is:

$$L(\lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

Take the natural logarithm to get the log-likelihood function, $\ell(\lambda)$:

$$\ell(\lambda) = \ln L(\lambda) = \sum_{i=1}^n (-\lambda + x_i \ln \lambda - \ln(x_i!))$$

$$\ell(\lambda) = -n\lambda + \ln(\lambda) \sum_{i=1}^n x_i - \sum_{i=1}^n \ln(x_i!)$$

Take the derivative with respect to λ and set it to 0 to find the maximum:

$$\begin{aligned} \frac{d}{d\lambda} \ell(\lambda) &= -n + \frac{\sum_{i=1}^n x_i}{\lambda} = 0 \\ n &= \frac{\sum_{i=1}^n x_i}{\lambda} \implies \hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x} \end{aligned}$$

The MLE for λ is the sample mean \bar{x} .

11. **Quiz in two parts**

An instructor has given a short quiz consisting of two parts. For a randomly selected student, let X be the number of points earned on the first part and Y be the number of points earned on the second part. Suppose that the joint pmf of X and Y is given by the accompanying table:

		y			
		0	5	10	15
x	0	0.02	0.06	0.02	0.10
	5	0.04	0.15	0.20	0.10
	10	0.01	0.15	0.14	0.01

a) Compute the marginal probability distributions of X and Y .

Solution: Summing across rows for X and columns for Y :

- $P(X = 0) = 0.02 + 0.06 + 0.02 + 0.10 = 0.20$
- $P(X = 5) = 0.04 + 0.15 + 0.20 + 0.10 = 0.49$
- $P(X = 10) = 0.01 + 0.15 + 0.14 + 0.01 = 0.31$

- $P(Y = 0) = 0.02 + 0.04 + 0.01 = 0.07$
- $P(Y = 5) = 0.06 + 0.15 + 0.15 = 0.36$
- $P(Y = 10) = 0.02 + 0.20 + 0.14 = 0.36$

- $P(Y = 15) = 0.10 + 0.10 + 0.01 = 0.21$

b) Compute the conditional probability mass function of Y given that $X = 10$.

Solution: $P(Y = y|X = 10) = \frac{P(X=10,Y=y)}{P(X=10)}$

- $P(Y = 0|X = 10) = \frac{0.01}{0.31} \approx 0.0323$
- $P(Y = 5|X = 10) = \frac{0.15}{0.31} \approx 0.4839$
- $P(Y = 10|X = 10) = \frac{0.14}{0.31} \approx 0.4516$
- $P(Y = 15|X = 10) = \frac{0.01}{0.31} \approx 0.0323$

c) Compute the conditional mean and variance of X given that $Y = 5$.

Solution: First, find the conditional pmf of X given $Y = 5$: $P(X = x|Y = 5) = \frac{P(X=x,Y=5)}{P(Y=5)} = \frac{P(x,5)}{0.36}$

- $P(X = 0|Y = 5) = \frac{0.06}{0.36} = \frac{1}{6}$
- $P(X = 5|Y = 5) = \frac{0.15}{0.36} = \frac{5}{12}$
- $P(X = 10|Y = 5) = \frac{0.15}{0.36} = \frac{5}{12}$

Conditional Mean $E[X|Y = 5]$:

$$E[X|Y = 5] = 0 \left(\frac{1}{6}\right) + 5 \left(\frac{5}{12}\right) + 10 \left(\frac{5}{12}\right) = \frac{25}{12} + \frac{50}{12} = \frac{75}{12} = 6.25$$

Conditional Variance $Var(X|Y = 5) = E[X^2|Y = 5] - (E[X|Y = 5])^2$:

$$E[X^2|Y = 5] = 0^2 \left(\frac{1}{6}\right) + 5^2 \left(\frac{5}{12}\right) + 10^2 \left(\frac{5}{12}\right) = \frac{125}{12} + \frac{500}{12} = \frac{625}{12}$$

$$Var(X|Y = 5) = \frac{625}{12} - (6.25)^2 = 52.0833 - 39.0625 = 13.0208$$

d) Are X and Y independent; justify your conclusion numerically.

Solution: No, they are not. If they were, $P(X, Y) = P(X)P(Y)$ for all x, y . Let's check $X = 0, Y = 0$: $P(X = 0, Y = 0) = 0.02$ $P(X = 0)P(Y = 0) = (0.20)(0.07) = 0.014$ Since $0.02 \neq 0.014$, they are not independent.

12. Voters

In a certain town, 40% of the eligible voters prefer candidate A, 10% prefer candidate B, and the remaining 50% have no preference. You randomly sample 12 eligible voters.

a) What is the probability that 6 will prefer candidate A, 2 will prefer candidate B, and the remaining 4 will have no preference?

Solution: Using the multinomial probability formula:

$$P(A = 6, B = 2, N = 4) = \frac{12!}{6!2!4!}(0.4)^6(0.1)^2(0.5)^4$$

$$P = (13860)(0.004096)(0.01)(0.0625) \approx 0.0355$$

b) What is the marginal distribution for the number of voters who prefer candidate A in your sample?

Solution: The marginal distribution of a single category in a multinomial distribution follows a Binomial distribution. Let N_A be the number of voters who prefer A. $N_A \sim \text{Binomial}(n = 12, p = 0.4)$. The pmf is:

$$P(N_A = k) = \binom{12}{k} (0.4)^k (0.6)^{12-k} \quad \text{for } k = 0, 1, \dots, 12$$

c) What is the marginal joint pmf for the number of voters who prefer candidates A, B in your sample?

Solution: Let a be the number who prefer A, and b be the number who prefer B. The number with no preference is strictly $12 - a - b$.

$$P(N_A = a, N_B = b) = \frac{12!}{a!b!(12 - a - b)!} (0.4)^a (0.1)^b (0.5)^{12-a-b}$$

where $a, b \geq 0$ and $a + b \leq 12$.

d) What is the conditional joint distribution for the number of voters who prefer candidates A, B in your sample given that 5 voters in the sample have no preference?

Solution: If 5 voters have no preference, then exactly 7 voters prefer either A or B. This effectively becomes a Binomial distribution among the remaining $n = 7$ voters, where the probability of preferring A is normalized against preferring A or B: $p'_A = \frac{p_A}{p_A + p_B} = \frac{0.4}{0.4 + 0.1} = \frac{0.4}{0.5} = 0.8$. Therefore, given $N_{\text{none}} = 5$, the number of voters preferring A follows a Binomial($n = 7, p = 0.8$) distribution. The conditional pmf for $A = a$ and $B = b$ (where $b = 7 - a$) is:

$$P(N_A = a, N_B = b | N_{\text{none}} = 5) = \binom{7}{a} (0.8)^a (0.2)^{7-a}$$

for $a = 0, 1, \dots, 7$ and $b = 7 - a$.