HOMEWORK 7
PROBABILITY & STATISTICS *

10-606 MATHEMATICAL FOUNDATIONS FOR MACHINE LEARNING

START HERE: Instructions

- **Collaboration Policy**: Please read the collaboration policy in the syllabus.
- **Late Submission Policy:** See the late submission policy in the syllabus.
- **Submitting your work:** You will use Gradescope to submit answers to all questions.
  - **Written:** For written problems such as short answer, multiple choice, derivations, proofs, or plots, please use the provided template. Submissions can be handwritten onto the template, but should be labeled and clearly legible. If your writing is not legible, you will not be awarded marks. Alternatively, submissions can be written in LaTeX. Each derivation/proof should be completed in the boxes provided. To receive full credit, you are responsible for ensuring that your submission contains exactly the same number of pages and the same alignment as our PDF template.
  - **Latex Template:** https://www.overleaf.com/read/rnyxwhkqhpgd

Question	Points
Expectation and Variance	8
Gaussian Likelihood	12
Total:	20

*Compiled on Monday 7th October, 2024 at 17:51

Instructions for Specific Problem Types

For “Select One” questions, please fill in the appropriate bubble completely:

Select One: Who taught this course?

● Matt Gormley

○ Marie Curie

○ Noam Chomsky

If you need to change your answer, you may cross out the previous answer and bubble in the new answer:

Select One: Who taught this course?

● Henry Chai

○ Marie Curie

~~●~~ Noam Chomsky

For “Select all that apply” questions, please fill in all appropriate squares completely:

Select all that apply: Which are scientists?

■ Stephen Hawking

■ Albert Einstein

■ Isaac Newton

□ I don’t know

Again, if you need to change your answer, you may cross out the previous answer(s) and bubble in the new answer(s):

Select all that apply: Which are scientists?

■ Stephen Hawking

■ Albert Einstein

■ Isaac Newton

□ I don’t know

For questions where you must fill in a blank, please make sure your final answer is fully included in the given space. You may cross out answers or parts of answers, but the final answer must still be within the given space.

Fill in the blank: What is the course number?

10-606	10-60~~6~~7

1 Expectation and Variance (8 points)

Consider the random variable $X = \binom{X_1}{X_2}$ whose distribution places equal probability on each of the following vectors $X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}$: $\binom{0}{4}, \binom{1}{3}, \binom{1}{5}, \binom{0}{2}$. (Notation: we are using subscripts to mean indices into a vector e.g. X_1, and superscripts to distinguish different vectors i.e. different possible values for X e.g. $X^{(2)}$)

1. (2 points) What is $E[X_1]$?

2. (3 points) What is $E[X_2 \mid X_1 = 1]$?

   Hint: We saw in the class that the expectation of a random variable Z is given as

   $$E[Z] = \sum_{z \in \mathcal{Z}} z \cdot P(Z = z)$$

   where $\mathcal{Z}$ is the set of possible values that Z can take. Conditional expectation is similar to the definition of expectation covered in the class, except that the probabilities used in the summation are conditional probabilities. Thus, the conditional expectation of Z given that $Y = y$ is given by

   $$E[Z \mid Y = y] = \sum_{z \in \mathcal{Z}} z \cdot P(Z = z \mid Y = y)$$

   .

3. (3 points) In the same distribution, what is $Var(X_2 \mid X_1 = 1)$?

   Hint: We saw in the class that the variance of a random variable Z is given as

   $$Var[Z] = E[(Z - E[Z])^2] = \sum_{z \in \mathcal{Z}} (z - E[Z])^2 \cdot P(Z = z)$$

   where $\mathcal{Z}$ is the set of possible values that Z can take. Conditional variance is similar to the definition of variance covered in the class, except that the probabilities used in the summation are conditional probabilities. Thus, the conditional variance of Z given that $Y = y$ is given by

   $$Var[Z \mid Y = y] = \sum_{z \in \mathcal{Z}} (z - E[Z])^2 \cdot P(Z = z \mid Y = y)$$

   .

2 Gaussian Likelihood (12 points)

Consider a Gaussian random variable $\mathbf{x} \sim \mathcal{N}(\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})$, where $\mathbf{x} \in \mathbb{R}^d$. Furthermore, we have

$$\mathbf{y} = A\mathbf{x} + \mathbf{b} + \mathbf{w},$$

where $\mathbf{y} \in \mathbb{R}^c$, $A \in \mathbb{R}^{c \times d}$, $\mathbf{b} \in \mathbb{R}^c$, and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, Q)$ is independent Gaussian noise. "Independent" implies that $\mathbf{x}$ and $\mathbf{w}$ are independent random variables and that Q is diagonal.

1. (4 points) Write down the likelihood $p(\mathbf{y}|\mathbf{x})$.

2. (8 points) The distribution $p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x}$ is Gaussian. Compute the mean $\mu_{\mathbf{y}}$ and the covariance $\Sigma_{\mathbf{y}}$. Show your work.

3 Collaboration Questions

After you have completed all other components of this assignment, report your answers to these questions regarding the collaboration policy. Details of the policy can be found in the syllabus.

1. Did you receive any help whatsoever from anyone in solving this assignment? If so, include full details.
2. Did you give any help whatsoever to anyone in solving this assignment? If so, include full details.
3. Did you find or come across code that implements any part of this assignment? If so, include full details.

Your Answer