AIS5103 homework 1 (due 1 Sep 2026)

Work out your solutions, preferably typing them neatly. Steps and reasoning are required, not just the final answers. It is best to work on the problems without the help of AI or the internet, as these questions are intended to help you for the midterm and final, which are closed-book. Submit your work to Canvas at "Assignment" in PDF format.

1. Read Bishop's book, "Deep Learning", Chapter 1.3, pages 16-22. Record your reading by taking notes and stating the main points of this reading. An AI-generated summary is not accepted. You must do the reading and write your report yourself.

2. Automatic differentiation in PyTorch, autograd. A tensor defined in PyTorch has the advantage that the partial derivatives of its elements with respect to a scalar can be computed automatically, provided you have the code for computing the scalar from the tensor elements. No implementation of numerical differentiation is needed, and the computation of the derivative is exact. Consider the function,

$$f(x) = \frac{e^x}{\sqrt{\sin^3 x + \cos^3 x}}$$

   Program this function in the interval for $0 \le x \le 1$ and also compute its derivative in two ways.

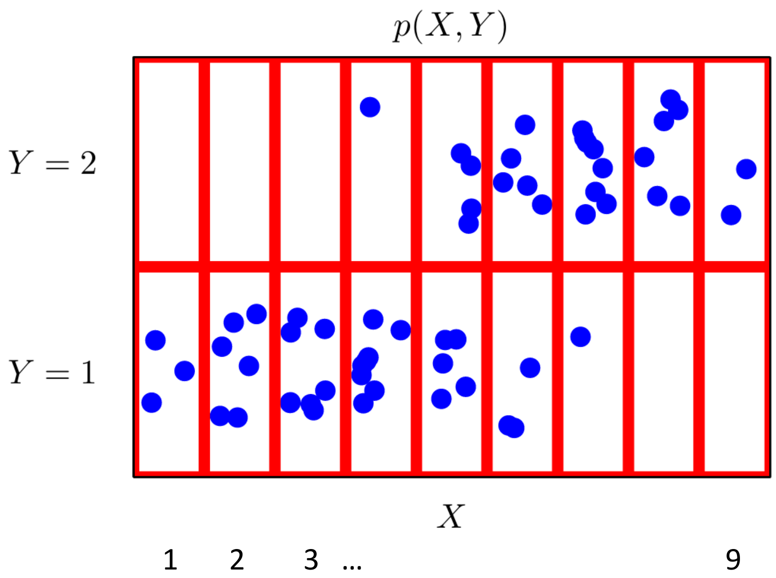
   a. By the PyTorch automatic differentiation (use the requires_grad=True option when defining the tensor, and for plotting, use the detach() method).
   b. By numerical finite difference. Examine the effect of step size.

   Compare the two results and their differences, and make appropriate plots. Submit a Jupyter notebook code in PDF as part of your overall submission of this question.

3. Consider Figure 2.5 on Bishop's book page 29, also shown below. There are two variables x and y. x = 1,2,…,9 (from left to right), y = 1, 2.
   a. Make a table of the joint distribution p(x, y), normalized properly. This will be used in the following questions.
   b. What is the marginal probability P(y=1)? What is the marginal probability P(x=1)?
   c. What is the conditional probability P(y=2|x=1)? What is the conditional probability P(x=1|y=1)?
   d. Compute E[x], and compute Var[x].
   e. Compute E_y [E_x[x|y]]. Prove in general the equality:

      E[x] = E_y [E_x[x|y]]

   Here E_x[x|y] denotes the expectation of x under the conditional distribution P(x|y), [This question clarifies the concepts of expectation E[ ], conditional expectation, variance Var[ ], joint distribution, marginal distribution, and conditional distribution.]



4. The Kullback-Leibler divergence between two probability distributions P and Q can be expressed as $D(P||Q) = \langle \ln \frac{P}{Q} \rangle_P$. The angular brackets denote the expectation $\langle f \rangle_P = E_P f$. Consider the die experiment in which we assume in theory (for Q) each of the six outcomes, 1 through 6, is equally probable. The actual probability P may be different.
    a. Show that if $P = Q$, then $D(P||Q) = 0$.
    b. Compute the Kullback-Leibler divergence if $P = (1,0,0,0,0,0)$, i.e., the probability of getting 1 is 1, and all others are 0.
    c. Prove that $D(P||Q) \geq 0$, i.e., the Kullback-Leibler divergence cannot be negative.

- The end -