City University of Hong Kong

CS5487 Machine Learning

Semester B 2021/22
Online Midterm
Time: 2 hours

1. The midterm has 5 pages including this page, consisting of 3 questions.
2. The following resources are allowed on the midterm:
   - You are allowed a cheat sheet that is **one** A4 page (**single-sided only**) handwritten with pen or pencil.
3. All other resources are not allowed, e.g., internet searches, classmates, textbooks.
4. Answer the questions on physical paper using pen or pencil.
   - Answer **ALL** questions.
   - Remember to write your **name, EID, and student number** at the top of each answer paper.
5. You should stay on Zoom during the entire exam time.
   - If you have any questions, use the private chat function in Zoom to message Antoni.
6. Midterm submission
   - Take pictures of your answer paper and submit it to the "Midterm Quiz" Canvas assignment. You may submit it as jpg/png/pdf.
   - *It is the student's responsibility to make sure that the captured images are legible. Illegible images will be graded as is, similar to illegible handwriting.*
   - If you have problems submitting to Canvas, then email your answer paper to Antoni (abchan@cityu.edu.hk).

Statement of Academic Honesty

Below is a **Statement of Academic Honesty**. Please read it.

I pledge that the answers in this exam are my own and that I will not seek or obtain an unfair advantage in producing these answers. Specifically,

- I will not plagiarize (copy without citation) from any source;
- I will not communicate or attempt to communicate with any other person during the exam; neither will I give or attempt to give assistance to another student taking the exam; and
- I will use only approved devices (e.g., calculators) and/or approved device models.
- I understand that any act of academic dishonesty can lead to disciplinary action.

I pledge to follow the Rules on Academic Honesty and understand that violations may led to severe penalties.

Name:
EID:
Student ID:
Signature:

(a) If you have not already, *copy the entire above statement of academic honesty to your answer sheet*. Fill in your name, EID, and student ID, and sign your signature to show that you agree with the statement and will follow its terms.

Problem 1   MLE for Exponential distribution [30 marks]

In this problem you will consider the maximum likelihood estimate (MLE) of the parameters of a *Exponential Distribution.*

The exponential distribution is a distribution of the time interval between consecutive events occurring in a Poisson process. A Poisson process is a process where events occur continuously at some constant average rate. For example, if we model a telephone switch as a Poisson process, then the time between incoming calls can be modeled as an exponential distribution.

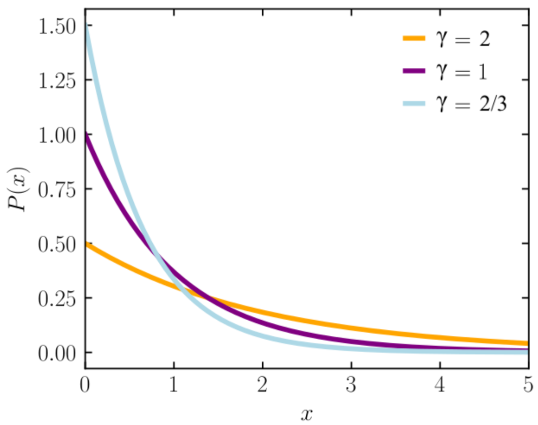
Formally, the exponential distribution is:

$$p(x|\gamma) = \begin{cases} \frac{1}{\gamma} e^{-x/\gamma}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{1}$$

where $\gamma > 0$ is the scale parameter or average interval length. The mean and variance of the exponential distribution are

$$E[x] = \gamma, \qquad \mathrm{var}(x) = \gamma^2. \tag{2}$$

Here is a plot of the exponential distribution for different values of γ:



Suppose we have a set of N samples of time intervals, $\mathcal{D} = \{x_1, \cdots, x_N\}$ with $x_i \geq 0$.

(a) [5 marks] Write down the log-likelihood of the data $\mathcal{D}$, i.e., $\log p(\mathcal{D}|\gamma)$.

(b) [5 marks] Write down optimization problem for maximum-likelihood estimation of the parameter γ.

(c) [15 marks] Derive the MLE for the parameter γ.

(d) [5 marks] What is the intuitive interpretation of the derived MLE for γ.

. . . .

Problem 2 MAP for Exponential distribution [35 marks]

In this problem you will compute the MAP estimate for exponential distribution in Problem 1. Let the prior distribution of γ be an Inverse Gamma distribution,

$$p(\gamma) = \text{InvGamma}(\gamma|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}(1/\gamma)^{\alpha+1}e^{-\beta/\gamma} \tag{3}$$

where $\alpha > 0$ and $\beta > 0$ are the shape and scale parameters of the Inverse Gamma prior (i.e., *hyperparameters*). $\Gamma(z)$ is the gamma function, which is an extension of factorial to positive real numbers, defined by

$$\Gamma(z) = \int_0^\infty u^{z-1}e^{-u}du. \tag{4}$$

The Gamma function obeys the relation $\Gamma(z+1) = z\Gamma(z)$. Furthermore, $\Gamma(1) = 1$, and $\Gamma(z+1) = z!$ when z is a non-negative integer.

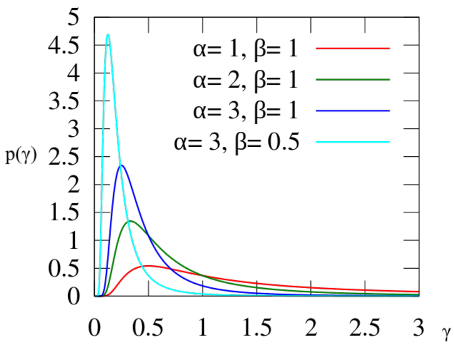
The mean, variance, and mode of $\text{InvGamma}(\alpha, \beta)$ are:

$$E[\gamma] = \frac{\beta}{\alpha - 1} \text{ for } \alpha > 1, \tag{5}$$

$$\text{var}(\gamma) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} \text{ for } \alpha > 2, \tag{6}$$

$$\text{mode}(\gamma) = \frac{\beta}{\alpha+1}. \tag{7}$$

The following plot shows some examples of the Inverse Gamma distribution for different parameters (α, β):



(a) [5 marks] Write down the optimization problem for MAP estimation of γ using an Inverse Gamma prior with known hyperparameters (α, β).

(b) [15 marks] Derive the MAP estimator for the parameter γ using the Inverse Gamma prior.

(c) [10 marks] Compare this MAP estimator with the ML estimator derived in Problem 1. What is the intuitive interpretation of the MAP solution with regards to the ML solution?

(d) [5 marks] What is the intuitive effect on the MAP estimate as the number of samples N increases?

....

Problem 3 Bayesian estimation for Exponential distribution [35 marks]

In this problem you will derive the Bayesian estimate for the Exponential distribution with Inverse Gamma prior, using the same setup as Problems 1 and 2.

(a) [5 marks] Write down the posterior distribution of the parameters, $p(\gamma|\mathcal{D})$, in terms of the prior and likelihood function.

(b) [15 marks] Derive the form of the posterior distribution $p(\gamma|\mathcal{D})$ in terms of the hyperparameters (α, β) and data $\mathcal{D}$.

(c) [5 marks] What is the intuitive interpretation of the derived Bayesian estimate in Problem 3(b), e.g., the mean of the posterior?

(d) [10 marks] What happens to the posterior distribution as the number of samples N increases?

....

— End of Midterm —