CS 6190: Probabilistic Machine Learning Spring 2022

Homework 0

Handed out: 10 Jan, 2022
Due: 11:59pm, 21 Jan, 2022

- You are welcome to talk to other members of the class about the homework. I am more concerned that you understand the underlying concepts. However, you should write down your own solution. Please keep the class collaboration policy in mind.
- Feel free discuss the homework with the instructor or the TAs.
- Your written solutions should be brief and clear. You need to show your work, not just the final answer, but you do *not* need to write it in gory detail. Your assignment should be **no more than 10 pages**. Every extra page will cost a point.
- Handwritten solutions will not be accepted.
- The homework is due by **midnight of the due date**. Please submit the homework on Canvas.

Warm up[100 points + 5 bonus]

1. [2 points] Given two events A and B, prove that

$$p(A \cup B) \le p(A) + p(B)$$
$$p(A \cap B) \le p(A), p(A \cap B) \le p(B)$$

   When does the equality hold?

2. [2 points] Let $\{A_1, \ldots, A_n\}$ be a collection of events. Show that

$$p(\cup_{i=1}^n A_i) \le \sum_{i=1}^n p(A_i).$$

   When does the equality hold? (Hint: induction)

3. [14 points] We use $\mathbb{E}(\cdot)$ and $\mathbb{V}(\cdot)$ to denote a random variable's mean (or expectation) and variance, respectively. Given two discrete random variables X and Y, where $X \in \{0, 1\}$ and $Y \in \{0, 1\}$. The joint probability $p(X, Y)$ is given in as follows:

	$Y = 0$	$Y = 1$
$X = 0$	$3/10$	$1/10$
$X = 1$	$2/10$	$4/10$

   (a) [10 points] Calculate the following distributions and statistics.

      i. the the marginal distributions $p(X)$ and $p(Y)$

ii. the conditional distributions $p(X|Y)$ and $p(Y|X)$

iii. $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\mathbb{V}(X)$, $\mathbb{V}(Y)$

iv. $\mathbb{E}(Y|X=0)$, $\mathbb{E}(Y|X=1)$, $\mathbb{V}(Y|X=0)$, $\mathbb{V}(Y|X=1)$

v. the covariance between X and Y

(b) [2 points] Are X and Y independent? Why?

(c) [2 points] When X is not assigned a specific value, are $\mathbb{E}(Y|X)$ and $\mathbb{V}(Y|X)$ still constant? Why?

4. [9 points] Assume a random variable X follows a standard normal distribution, i.e., $X \sim \mathcal{N}(X|0,1)$. Let $Y = e^{-X^2}$. Calculate the mean and variance of Y.

   (a) $\mathbb{E}(Y)$

   (b) $\mathbb{V}(Y)$

   (c) $\text{cov}(X,Y)$

5. [8 points] Derive the probability density functions of the following transformed random variables.

   (a) $X \sim \mathcal{N}(X|0,1)$ and $Y = X^3$.

   (b) $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} | \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix})$ and $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1/2 \\ -1/3 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$.

6. [10 points] Given two random variables X and Y, show that

   (a) $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$

   (b) $\mathbb{V}(Y) = \mathbb{E}(\mathbb{V}(Y|X)) + \mathbb{V}(\mathbb{E}(Y|X))$

   (Hints: using definition.)

7. [9 points] Given a logistic function, $f(\mathbf{x}) = 1/(1+\exp(-\mathbf{a}^\top \mathbf{x}))$ ($\mathbf{x}$ is a vector),

   (a) derive $\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}}$

   (b) derive $\frac{\mathrm{d}^2 f(\mathbf{x})}{\mathrm{d}\mathbf{x}^2}$, i.e., the Hessian matrix

   (c) show that $-f(\mathbf{x})$ is convex

   Note that $0 \le f(\mathbf{x}) \le 1$.

8. [10 points] Derive the convex conjugate for the following functions

   (a) $f(x) = -\log(x)$

   (b) $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}^{-1}\mathbf{x}$ where $\mathbf{A} \succ 0$

9. [20 points] Derive the (partial) gradient of the following functions. Note that bold small letters represent vectors, bold captial letters matrices, and non-bold letters just scalars.

   (a) $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$, derive $\frac{\partial f}{\partial \mathbf{x}}$

   (b) $f(\mathbf{x}) = \left(\mathbf{I} + \mathbf{x}\mathbf{x}^\top\right)^{-1} \mathbf{x}$, derive $\frac{\partial f}{\partial \alpha}$

   (c) $f(\mathbf{x}) = \log|\mathbf{K} + \alpha \mathbf{I}|$, where $|\cdot|$ means the determinant. Derive $\frac{\partial f}{\partial \sigma}$

   (d) $f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \log\left(\mathcal{N}(\mathbf{a}|\mathbf{A}\boldsymbol{\mu}, \mathbf{S}\boldsymbol{\Sigma}\mathbf{S}^\top)\right)$, derive $\frac{\partial f}{\partial \boldsymbol{\mu}}$ and $\frac{\partial f}{\partial \boldsymbol{\Sigma}}$,

   (e) $f(\boldsymbol{\Sigma}) = \log\left(\mathcal{N}(\mathbf{a}|\mathbf{b}, \mathbf{K} \otimes \boldsymbol{\Sigma})\right)$ where $\otimes$ is the Kronecker product (Hint: check Minka's notes).

10. [2 points] Given the multivariate Gaussian probability density,

$$p(\mathbf{x}|\boldsymbol{\mu}, \mathbf{\Sigma}) = |2\pi\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left(-(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right).$$

Show that the density function achieves the maximum when $\mathbf{x} = \boldsymbol{\mu}$.

11. [5 points] Show that

$$\int \exp(-\frac{1}{2\sigma^2}x^2)\mathrm{d}x = \sqrt{2\pi\sigma^2}.$$

Note that this is about how the normalization constant of the Gaussian density is obtained. Hint: consider its square and use double integral.

12. [5 points] The gamma function is defined as

$$\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}\mathrm{d}u.$$

Show that $\Gamma(1) = 1$ and $\Gamma(x+1) = x\Gamma(x)$. Hint: using integral by parts.

13. [2 points] By using Jensen's inequality with $f(x) = \log(x)$, show that for any collection of positive numbers $\{x_1, \ldots, x_N\}$,

$$\frac{1}{N}\sum_{n=1}^N x_n \geq \left(\prod_{n=1}^N x_n\right)^{\frac{1}{N}}.$$

14. [2 points] Given two probability density functions $p(\mathbf{x})$ and $q(\mathbf{x})$, show that

$$\int p(\mathbf{x})\log\frac{p(\mathbf{x})}{q(\mathbf{x})}\mathrm{d}\mathbf{x} \geq 0.$$

15. [**Bonus**][5 points] Show that for any square matrix $\mathbf{X} \succ 0$, $\log|\mathbf{X}|$ is concave to $\mathbf{X}$.