

MATH FOUNDATION OF DATA SCIENCE HOMEWORK 2

>>NAME: <<
>>WISC ID: <<

Instructions: Please complete the homework in LaTeX, produce a pdf, and submit it on time as indicated on Moodle.

1 Probability Theory Basics

Q1 [5 points] Use the axioms of probability to prove that

$$\forall A \in \mathcal{F}, P(A^c) = 1 - P(A).$$

[Solution:](#)

Q2 [5 points] Suppose events A, B, C are mutually independent. Prove that $A \cup B$ and C are independent.

[Solution:](#)

Q3 [10 points] Suppose a sequence of events $\{A_n : n \geq 1\}$ is a monotonically increasing sequence, i.e., $A_n \subseteq A_{n+1}, \forall n \geq 1$. Define $\lim_{n \rightarrow \infty} A_n = \cup_{i=1}^{\infty} A_i$. Prove the continuity theorem for probability using the axioms:

$$\lim_{n \rightarrow \infty} P(A_n) = P(\lim_{n \rightarrow \infty} A_n).$$

Solution:

Q4 [10 points] The moment generating function (MGF) of a random variable X is defined as:

$$\psi(\lambda) := \mathbb{E}e^{\lambda X} = \int_{-\infty}^{+\infty} e^{\lambda x} dF_X(x)$$

if the right-hand side exists. Prove that if the k -th moment of X exists, then

$$\mathbb{E}X^k = \psi^{(k)}(0).$$

Solution:

2 Statistical Inference and Learning

Q5 [10 points] Let X_1, \dots, X_n be a sample from a geometric distribution with unknown parameter θ , i.e., its density is $p(x; \theta) = (1 - \theta)^{x-1} \theta$. What is the maximum likelihood estimator (MLE) of θ ?

[Solution:](#)

Q6 [10 points] Let X be a continuous random variable with the following PDF:

$$p(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Also, suppose given X, Y has distribution:

$$P(y|x) = x(1-x)^{y-1}, \quad \text{for } y = 1, 2, 3, \dots \quad (2)$$

What is the MAP estimate of X given $Y = 2$?

Solution:

Q7 [10 points] The infection rate of some bacteria in the whole population is 0.02. The result of testing is either positive or negative, but the result may be inaccurate. Suppose:

$$\begin{aligned}P(\text{negative}|\text{infected}) &= 0.01, \\P(\text{negative}|\text{noninfected}) &= 0.97.\end{aligned}$$

If one is tested positive, what is the probability of him/her being infected? (The solution should be in fraction form, i.e., p/q for two integers p, q .)

[Solution:](#)

Q8 [10 points] For linear regression, consider the following statistical model for the data (x, y) :

$$y = \theta^\top x + \epsilon$$

where ϵ has a Laplace distribution

$$p(\epsilon) = \frac{1}{2b} \exp\left(-\frac{|\epsilon|}{b}\right)$$

for some $b > 0$. Also assume a Laplace prior for θ : each dimension θ_i is independent and

$$p(\theta_i) = \frac{1}{2b_\theta} \exp\left(-\frac{|\theta_i|}{b_\theta}\right)$$

for some $b_\theta > 0$.

Given a set of data points $\{(x_i, y_i)\}_{i=1}^n$, use MAP to derive the training objective for θ .

Solution:

Q9 [10 points] Let X_1, \dots, X_n be a random sample from an exponential population with pdf:

$$p(x|\theta) = \begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & x < \theta, \end{cases}$$

where $-\infty < \theta < \infty$. Consider testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where θ_0 is a fixed value. Compute the likelihood ratio test statistic $\lambda(x_1, \dots, x_n)$.

Solution:

3 Information Theory

Q10 [10 points] The conditional mutual information of random variables X and Y given Z is defined by

$$I(X; Y|Z) = H(X|Z) - H(X|Y, Z) = \mathbb{E}_{p(x,y,z)} \log \frac{p(X, Y|Z)}{p(X|Z)p(Y|Z)}.$$

For joint probability mass functions $p(x, y)$ and $q(x, y)$, the conditional relative entropy $D(p(y|x)||q(y|x))$ is the average of the relative entropies between the conditional probability mass functions $p(y|x)$ and $q(y|x)$ averaged over the probability mass function $p(x)$. More precisely,

$$D(p(y|x); q(y|x)) = \sum_x p(x) \sum_y p(y|x) \log \frac{p(y|x)}{q(y|x)} = \mathbb{E}_{p(x,y)} \log \frac{p(Y|X)}{q(Y|X)}.$$

Prove:

(1) Chain rule for entropy: let X_1, \dots, X_n be drawn according to $p(x_1, x_2, \dots, x_n)$. Then

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1).$$

(2) Chain rule for mutual information:

$$I(X_1, X_2, \dots, X_n; Y) = \sum_{i=1}^n I(X_i; Y | X_{i-1}, \dots, X_1).$$

(3) Chain rule for relative entropy:

$$D(p(x, y)||q(x, y)) = D(p(x)||q(x)) + D(p(y|x)||q(y|x)).$$

Solution:

4 Concentration

Q11 [10 points] Imagine we have an algorithm for solving some decision problem (e.g. is a given number p a prime?). Suppose the algorithm makes a decision at random and returns the correct answer with probability $1/2 + \delta$ with some $\delta > 0$, which is just a bit better than a random guess. To improve the performance, we run the algorithm N times and take the majority vote. Show that, for any $\epsilon \in (0, 1)$, the answer is correct with probability at least $1 - \delta$, as long as

$$N \geq \frac{1}{2\delta^2} \ln \left(\frac{1}{\epsilon} \right).$$

Solution: