

EN.601.448/648 Computational genomics: Problem set 0

LuchaoQi (lqi9)

Instructions

We have provided this L^AT_EX document for turning in Problem set 0. We give you one or more boxes to answer each question. The question to answer for each box will be noted in the title of the box.

Other than your name, do not type anything outside the boxes. Leave the rest of the document unchanged.

For written answers, replace the `\TextRequired (Place Answer Here)` command with your answer. For the following example *Question 0.1*, you would place your answer where `\TextRequired (Place Answer Here)` is located,

Question 0.1**Place Answer Here**

Do not change the height or title of the box. If your text goes beyond the box boundary, it will be cut off. We have given sufficient space for each answer, so please condense your answer if it overflows. The height of the box is an upper bound on the amount of text required to answer the question - many answers can be answered in a fraction of the space. Do not add text outside of the boxes. We will not read it.

For True/False or Multiple Choice questions, place your answers within the defined table. To mark the box(es) corresponding to your answers, replace `\Unchecked (☐)` commands with the `\Checked (☒)` command. Do not make any other changes to the table. For example, in *Question 0.2*,

Question 0.2

- | | |
|-------------------------------------|---------------------|
| <input checked="" type="checkbox"/> | Logistic Regression |
| <input type="checkbox"/> | Perceptron |

For answers that require a single equation, we will provide a specific type of box, such as in the following example *Question 0.3*. Please type the equation where `\EquationRequired (Type Equation Here)` without adding any \$ signs or `\equation` commands. Do not put any additional text in this field.

Question 0.3

 $w =$

Type Equation Here

For answers that require multiple equations, such as a derivation, place all equations within the specified box. You may include text short explanations if you wish (as shown in *Question 0.4*). You can put the equations in any format you like (e.g. within $\$$ or $\$\$$, the `\equation` environment, the `\align` environment) as long as they stay within the box.

Question 0.4

$$x + 2$$

x is a real number

the following equation uses the variable y

$$y + 3$$

Do not change any formatting in this document, or we may be unable to grade your work. This includes but is not limited to the height of textboxes, font sizes, and the spacing of text and tables. Additionally, do not add text outside of the answer boxes. Entering your answers are the only changes allowed.

We strongly recommend you review your answers in the generated PDF to ensure they appear correct. We will grade what appears in the answer boxes in the submitted PDF, NOT the original latex file.

1. Expected value and likelihood (1 point)

Question 1.1

$$P(30 \text{ tosses all yield heads}) = \theta^{30} \approx 2.21 * 10^{-7}$$

Question 1.2

$$\mathbb{E}(\text{One toss}) = 0.4$$

Question 1.3

$$\mathbb{E}(\text{sum of 4 coin tosses}) = 4 * 0.4 = 1.6$$

Question 1.4 (a)

$$\text{Given } \theta = 0.5, \text{ Likelihood} = 0.5^7 * 0.5^3 \approx 9.77 * 10^{-4}$$

Question 1.4 (b)

$$\text{Given } \theta = 0.8, \text{ Likelihood} = 0.8^7 * 0.2^3 \approx 1.68 * 10^{-3}$$

Question 1.4 (c) Which model do you prefer

☐ $\theta = 0.5$

☒ $\theta = 0.8$

Question 1.4 (c) Justification

The likelihood is larger when $\theta = 0.8$, which means $\theta = 0.8$ is better supported given the data.

Question 1.5 (d) Is there better setting?

$$\begin{aligned}\mathcal{L}(\theta|x) &= \theta^7(1-\theta)^3 \\ \log \mathcal{L}(\theta|x) &= 7\log(\theta) + 3\log(1-\theta) \\ \frac{d \log \mathcal{L}(\theta|x)}{d\theta} &= \frac{7}{\theta} - \frac{3}{1-\theta} = 0 \\ \theta &= 0.7\end{aligned}$$

So we have MLE of θ if $\theta = 0.7$

2. Probability and independence (0.5 points)

Question 2.1 Prove

$$P(A|B) = \frac{P(AB)}{P(B)} > P(B) \quad (1)$$

With equation 1, we have:

$$P(B|A) = \frac{P(AB)}{P(A)} > P(B) \quad (2)$$

Question 2.2 (a.1) are these two events independent

☐ Yes

☒ No

Question 2.2 (a.2) Justification

Because

$$\begin{aligned} P(AB) &= 8/25 \\ P(A) * P(B) &= \frac{13}{25} * \frac{10}{25} \\ P(AB) &\neq P(A) * P(B) \end{aligned}$$

So they are not independent.

Question 2.2 (b)

$$P = 1 - P(B|A) = 1 - \frac{8}{13} = \frac{5}{13}$$

Question 2.2 (c). which is more likely

☐ Steel

☒ Wooden

3. Bayes' Theorem (0.5 points)

Question 3.1 Probability of actually have the disease (write out the derivation)

$$\begin{aligned} P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^C)P(D^C)} \\ &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^C)(1 - P(D))} \\ &= \frac{0.96 * \frac{1}{250000}}{0.96 * \frac{1}{250000} + 0.005 * (1 - \frac{1}{250000})} \\ &\approx 7.67 * 10^{-4} \end{aligned}$$

4. Gaussian data and likelihood

Question 4.1

$$\mathbb{E}[\sum_{i=1}^{10} x_i] = 10 * 0 = 0$$

Question 4.2

$$E[\sum_{i=1}^{10} x_i^2] = 10 * (1 + 0^2) = 10$$

Question 4.3

Actual ten values (3 signif digits): $1.16 * 10^0, -7.38 * 10^{-1}, -8.61 * 10^{-1}, 6.31 * 10^{-1}, -6.17 * 10^{-1}, 9.58 * 10^{-1}, -1.71 * 10^0, 1.31 * 10^0, 9.41 * 10^{-1}, 2.08 * 10^0$

Sum = 3.154

Sum of squares = 14.18

Question 4.4

$$p(x_1) = 0.020$$

$$p(x_2) = 0.304$$

$$p(x_3) = 0.275$$

Question 4.5

$$\text{Likelihood}(x_1, x_2, \dots, x_{10}) = 2.86 * 10^{-7}$$

Question 4.6

$$\log \text{Likelihood}(x_1, x_2, \dots, x_{10}) = -0.153$$

Question 4.6 Why might people prefer to work in log space?

Taking logs typically makes log likelihoods better behaved. Also, it's easier to get maximum log likelihood than maximum likelihood.

5. Linear algebra (0.5 points)

Question 5.1 A is orthonormal

$A^T A$ is $A^{-1} A = I$, I is identity matrix.

Question 5.2 (a) If A is invertible write the inverse of A

$A^{-1} =$

$$\begin{bmatrix} \frac{1}{3} & 0 & \frac{1}{3} \\ -\frac{1}{6} & \frac{1}{2} & -\frac{1}{6} \\ \frac{1}{6} & -\frac{1}{2} & \frac{7}{6} \end{bmatrix}$$

Question 5.2 (a) If A is not invertible write the justification

Place Answer Here

Question 5.2 (b) If A is invertible write the inverse of A

$A^{-1} =$

$$\begin{bmatrix} \text{Place Answer Here} & \text{Place Answer Here} & \text{Place Answer Here} \\ \text{Place Answer Here} & \text{Place Answer Here} & \text{Place Answer Here} \\ \text{Place Answer Here} & \text{Place Answer Here} & \text{Place Answer Here} \end{bmatrix}$$

Question 5.2 (b) If A is not invertible write the justification

The determinant of matrix is 0, so A is not invertible.

Question 5.3 Implement the function in Python

6. Matrix derivatives. (1 point)

Question 6.1 Matrix derivative (write out the derivation)

$$\begin{aligned} Q &= (Y - X\beta)^T(Y - X\beta) \\ &= Y^TY - 2\beta X^TY + \beta X^TX\beta \\ \frac{dQ}{d\beta} &= -2X^TY + 2X^TX\beta \end{aligned}$$

Question 6.2 Optimal β

When

$$\begin{aligned} \frac{dQ}{d\beta} &= 0 \\ X^TX\beta &= X^TY \end{aligned}$$

If the matrix X^TX is invertible, we have:

$$\beta = (X^TX)^{-1}X^TY$$

Question 6.3 Unique solutions in linear regression

$\text{rank}(X^T X) = \text{rank}(X^T X | X^T Y)$ means we have solutions for β .

Iff $N \geq K$ will the matrix be a full rank matrix and the number of equations is larger than the number of variables, which means solution is unique