

Written assignment

1. Let x and y be two independent random variables, show that:

(a) For any constants a and b , $E(ax + by) = aE(x) + bE(y)$

$$\begin{aligned} E(ax + by) &= \int \int (ax + by)f(x, y) dx dy \\ &= \int \int (ax + by)f(x)f(y) dx dy \\ &= \int \int axf(x)f(y) dx dy + \int \int byf(x)f(y) dx dy \\ &= a \int xf(x) dx + b \int yf(y) dy \\ &= aE(x) + bE(y) \end{aligned}$$

(b) For any constants a and b , $Var(ax + by) = a^2Var(x) + b^2Var(y)$

$$\begin{aligned} Var(aX + bY) &= \int \int (ax + by - E(ax + by))^2 f(x, y) dx dy \\ &= \int \int a^2(x - E(x))^2 + b^2(y - E(y))^2 + 2ab(x - E(x))(y - E(y))f(x)f(y) dx dy \\ &= a^2Var(x) + b^2Var(y) + 2 \int a(x - E(x))f(x)b \int (y - E(y))f(y) dx dy \\ &= a^2Var(x) + b^2Var(y) \end{aligned}$$

(c) $Cov(x, y) = 0$

$$\begin{aligned} Cov(x, y) &= \int \int (x - E(x))(y - E(y))f(x, y) dx dy \\ &= \int (x - E(x))f(x) \int (y - E(y))f(y) dy dx \\ &= \int (x - E(x))f(x) \left(\int yf(y) dy - \int E(y)f(y) dy \right) dx \\ &= \int (x - E(x))f(x)(E(y) - E(y)) dx \\ &= 0 \end{aligned}$$

2. Consider two coins, one is fair and one is not. The unfair coin has a 1/10 probability for head. Now you close your eyes and pick a random coin (each coin has a 50% probability being picked), and toss it twice. Answer the following question.

(a) What is the probability of the first toss being head? Let H denote the event of getting a head. Let A and B denote the event that the fair and unfair coin was chosen respectively.

$$p(H) = p(H, A) + p(H, B) \tag{1}$$

$$= p(H|A)p(A) + p(H|B)p(B) \tag{2}$$

$$= 0.5 * 0.5 + 0.1 * 0.5 \tag{3}$$

$$= 0.3 \tag{4}$$

(b) If both tosses are heads, what is the probability the fair coin is chosen?

$$p(A|HH) = p(A, HH)/p(HH) \quad (5)$$

$$= p(HH|A)p(A)/P(HH) \quad (6)$$

$$= 0.25 * 0.5 / 0.13 \quad (7)$$

$$= 25/26 \quad (8)$$

$$p(HH) = p(HH|A)p(A) + p(HH|B)p(B) \quad (10)$$

$$= 0.25 * 0.5 + 0.01 * 0.5 \quad (11)$$

$$= 0.13 \quad (12)$$

3. We have three boxes colored Red, Blue and Green respectively. There are 3 apples and 6 oranges in the red box, 3 apples and 3 oranges in the green box, and 5 apples and 3 oranges in the blue box. Now we randomly select one of the boxes (with equal probability) and grab a fruit from the box. What is the probability that it is an apple? If the fruit that we got is an orange, what is the probability that we selected the green box?

$$p(a) = p(a|B)p(B) + p(a|G)p(G) + p(a|R)p(R) = \frac{5}{8} * \frac{1}{3} + \frac{1}{2} * \frac{1}{3} + \frac{1}{3} * \frac{1}{3} = \frac{35}{72}$$

$$p(G|o) = \frac{p(o|G)p(G)}{p(o)} = \frac{\frac{1}{2} * \frac{1}{3}}{1 - 35/72} = \frac{1}{6} * \frac{72}{37} = \frac{12}{37}$$

4. (Weighted linear regression) In class when discussing linear regression, we assume that the Gaussian noise is independently identically distributed. Now we assume the noises $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ are independent but each $\epsilon_m \sim N(0, \sigma_m^2)$, i.e., it has its own distinct variance.
- (a) Write down the log likelihood function of \mathbf{w} .
 - (b) Show that maximizing the log likelihood is equivalent to minimizing a weighted least square loss function $J(\mathbf{W}) = \sum_{m=1}^n a_m (\mathbf{w}^T \mathbf{x}_m - y_m)^2$, and express each a_m in terms of ϵ_m .
 - (c) Derive a batch gradient descent algorithm for optimizing this objective.
 - (d) Derive a closed form solution to this optimization problem.

$$\begin{aligned} \log p(D|M) &= \sum_{i=1}^N \log N(y_i | \mathbf{x}_i^T \mathbf{w}, \sigma_i^2) \\ &= \sum_{i=1}^N \log \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2\sigma_i^2}(y_i - \mathbf{x}_i^T \mathbf{w})^2} \\ &= \sum_{i=1}^N \log \frac{1}{\sqrt{2\pi}\sigma_i} - \sum_{i=1}^N \frac{1}{2\sigma_i^2} (y_i - \mathbf{x}_i^T \mathbf{w})^2 \end{aligned}$$

Maximizing the log likelihood is equivalent to minimizing the second term in the above equation, which can be represented as:

$$J(\mathbf{w}) = \sum_i a_i (y_i - \mathbf{x}_i^T \mathbf{w})^2$$

where $a_i = \frac{1}{2\sigma_i^2}$.

Let $\mathbf{y} = y_1, y_2, \dots, y_N^T$, \mathbf{X} be the data matrix whose rows correspond to training examples, and A be a diagonal matrix with $A(i, i) = a_i$. Rewriting the objective in matrix form:

$$\begin{aligned} J(\mathbf{w}) &= (\mathbf{y} - \mathbf{X}\mathbf{w})^T A (\mathbf{y} - \mathbf{X}\mathbf{w}) \\ &= \mathbf{y}^T A \mathbf{y} - 2\mathbf{w}^T \mathbf{X}^T A \mathbf{y} + \mathbf{w}^T \mathbf{X}^T A \mathbf{X} \mathbf{w} \end{aligned}$$

Take the gradient of J and setting it to zero:

$$\nabla_{\mathbf{w}} J = 2\mathbf{X}^T A \mathbf{X} \mathbf{w} - 2\mathbf{X}^T A \mathbf{y} = 0$$

$$2\mathbf{X}^T A \mathbf{X} \mathbf{w} = 2\mathbf{X}^T A \mathbf{y}$$

$$\mathbf{w}^* = (\mathbf{X}^T A \mathbf{X})^{-1} \mathbf{X}^T A \mathbf{y}$$

5. (Maximum likelihood estimation.) A DNA sequence is formed using four bases Adenine(A), Cytosine(C), Guanine(G), and Thymine(T). We are interested in estimating the probability of each base appearing in a DNA sequence. Note that here we are treating each DNA position's base as a random variable x following a categorical distribution of 4 values and assume the different sequences are generated by repeatedly sampling from this distribution. This distribution has 4 parameters, which we denote as p_a, p_c, p_g , and p_t here. Given a collection of DNA sequences, please derive a maximum likelihood estimation for p_a, p_c, p_g , and p_t .

Helpful Hint: the probability mass function for the discrete random variable can be written compactly as

$$p(x) = \prod_{s=a,c,g,t} p_s^{I(x=s)}$$

Here $I(x = s)$ is an indicator function, and takes value 1 if x is equal to s , and 0 otherwise.

We can think of x as the outcome of rolling a 4-sided die. To simplify the notation, we will use index 1, 2, 3, 4 to replace the four letters. The parameters are $\mathbf{p} = [p_1, p_2, p_3, p_4]$, subject to the constraint that $\sum_{i=1}^4 p_i = 1$.

Given a collection of DNA sequences, we could concatenate them into one long sequence of length n : $[x_1, x_2, \dots, x_n]$.

Because we consider each position as independent, the log likelihood function can be written as:

$$\begin{aligned} l(\mathbf{p}) &= \sum_{m=1}^n \log \prod_{i=1}^4 p_i^{I(x_m=i)} \\ &= \sum_{i=1}^4 \log p_i \sum_{m=1}^n I(x_m = i) \\ &= \sum_{i=1}^4 n_i \log p_i \end{aligned}$$

To take the constraints into consideration, we form the lagrangian:

$$l(\mathbf{p}, \lambda) = \sum_{i=1}^4 n_i \log p_i - \lambda(1 - \sum_{i=1}^4 p_i)$$

To find the optimizing parameters, we take the partial derivative of l wrt each parameter p_i , and setting it to zero:

$$\frac{\partial}{\partial p_i} l = \frac{n_i}{p_i} - \lambda = 0 \text{ for } i = 1, \dots, 4$$

Using the constraints $\sum_{i=1}^4 p_i = 1$, we arrive at the conclusion that $\lambda = n$. This leads to the final solution:

$$p_i = \frac{n_i}{n}, \text{ for } i = 1, \dots, 4$$