CSCI 5622 Fall 2023–HW0

Luk Letif

September 11, 2023

1 Probability and statistics

1.

(a)

$$p(\text{ all 20 students are from different countries}) = \frac{150}{150} \cdot \frac{149}{150} \cdot \frac{148}{150} \cdots \frac{131}{150} \approx 0.724$$

Therefore

$$p(\text{at least two students come from the same country}) =$$
$$1 - p(\text{ all 20 students are from different countries}) = 1 - 0.724 = 0.276$$

(b)

From part (a), we know

$$p(\text{at least two students come from the same country})$$
$$= 1 - p(\text{ all class students are from different countries})$$
$$= 1 - \frac{150}{150} \cdot \frac{149}{150} \cdot \frac{148}{150} \cdots \frac{150-n+1}{150}$$

Hence

$$1 - 1 \cdot \frac{149}{150} \cdot \frac{148}{150} \cdot \ldots \cdot \frac{150-n+1}{150} \geq 1 - 0.95$$
$$\Rightarrow n \geq 30$$

Class has to have 30 student so that there is .95 chance that two or more students come from the same country.

2.

We know that random variables X and Y are conditionally independent given $Y = y$ if and only if the conditional probability distribution of X given $Y = y$ is equal to the marginal probability distribution of X[1].

Let us check Conditional Probability Distribution of X given $Y = 2$:

$$P(X = 1|Y = 1) = \frac{1}{20}$$

$$P(X = 2|Y = 1) = \frac{2}{20}$$

$$P(X = 3|Y = 1) = \frac{2}{20}$$

[1] Any textbook

$$P(X = 4|Y = 1) = 0$$

Where Marginal Probability Distribution of X:

$$P(X = 1) = \frac{3}{20}$$

$$P(X = 2) = \frac{8}{20}$$

$$P(X = 3) = \frac{6}{20}$$

$$P(X = 4) = \frac{3}{20}$$

The conditional probability distribution of not equal to the marginal probability distribution. Therefore, X and Y are not independent.

3.

$$f_X(x) = \begin{cases} \frac{1}{10}, & 0 \leq x \leq 5 \\ c, & 6 < x \leq 7 \\ 0, & \text{otherwise} \end{cases}$$

To find c, we need

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

That is:

$$\int_0^5 \frac{1}{10}\, dx + \int_6^7 c\, dx = 1$$

Now, solve for c:

$$\frac{1}{10} x|_0^5 + cx|_6^7 = 1$$

$$\frac{1}{10}\,[5 - 0] + c\,[6 - 5] = 1$$

We get

$$c = \frac{1}{2}$$

4.

$$E[Y] = E[X + N] = E[X] + E[N]$$

where $E[X] = \dfrac{-2+2}{2} = 0$ and $E[N] = \dfrac{-1+1}{2} = 0$.
Therefore,

$$E[Y] = 0$$

5.

(a)

First, let's calculate $Cov(X, Y)$ using the definition:

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

We are given that $E(Y|X = x) = x$, which means the conditional expectation of Y given X is x. We can rewrite this as:

$$E(Y|X) = X$$

Therefore:

$$Cov(X, Y) = E[E[(X - E[X])(X - E[X])|X]] = E[E[X^2 - 2XE[X] + E[X]^2|X]]$$

Using the iterative expectation, we get:

$$Cov(X, Y) = E[X^2 - 2XE[X] + E[X]^2]$$

That is:

$$Cov(X, Y) = E[X^2 - 2XE[X] + E[X]^2] = E[(X - E[X])^2] = Var(X)$$

(b)

Since

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

we have X, y are independent, meaning $E[XY] = E[X]E[y]$, therefore:

$$Cov(X, Y) = E[XY] - E[X]E[Y] = 0.$$

6.

Given $M_n = \dfrac{\sum_{i=1}^{n} X_i}{n}$, we have:

$$E[M_n] = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} p = \frac{n}{n}p = p$$

Next, given $Var(\text{Bernoulli}) = p(1-p)$, we get:

$$Var(M_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}(np(1-p)) = \frac{p(1-p)}{n}$$

(b)

Using Chebyshev's inequality:

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

with finite non-zero variance μ^2 (and thus finite expected value σ). We get:

$$P(|M_n - p| \geq \epsilon) \leq \frac{Var(M_n)}{\epsilon^2} = \frac{\frac{p(1-p)}{n}}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2}$$

Therefore, our estimate M_n is more than ϵ away is bounded by $\frac{p(1-p)}{n\epsilon^2}$.

2 Linear algebra and calculus

1.

$$\mathbf{A} = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix}$$

(a)

$$\mathbf{Ac} = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \end{bmatrix} \text{does not exist, due to dimension not matched}$$

Unless you mean

$$\mathbf{Ac}^T = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$

(b)

$$\mathbf{Ax} = \mathbf{b}$$

Can be written as:

$$\begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$$

Using Gaussian elimination, we get:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \begin{bmatrix} -1 & 1 & 0 \\ -2 & 0.5 & 1 \\ 4 & -2.5 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 5.5 \end{bmatrix}$$

2.

With all the assumption about the matrix, we need multiply A^{-1} on both side of

$$AXA = Y - C$$

That is

$$A^{-1}(AXA)A^{-1} = A^{-1}(Y - C)A^{-1}$$

Since $A^{-1}A = AA^{-1} = I$ We get

$$X = A^{-1}(Y - C)A^{-1}$$

3.

(a)

With chain rule, we have

$$\frac{dL}{db_1} = \frac{dL}{dp} \cdot \frac{dp}{dz} \cdot \frac{dz}{db_1} = 2(y - p) \cdot \frac{2}{1 + e^{-z}} \left(1 - \frac{2}{1 + e^{-z}}\right) \cdot x$$

Where:

$$\begin{cases} \frac{dL}{dp} = 2(y - p) \\ \frac{dp}{dz} = \frac{2}{1+e^{-z}} \left(1 - \frac{2}{1+e^{-z}}\right) \\ \frac{dz}{db_1} = x \end{cases}$$

Therefore:
$$\frac{dL}{db_1} = 2(y-p)\cdot\frac{2}{1+e^{-z}}\left(1-\frac{2}{1+e^{-z}}\right)\cdot x = \frac{2x(y-p)e^{-z}}{(1+e^{-z})^2}$$

(b)

From part (a), we have:
$$\frac{dL}{db_1} = 2(y-p)\cdot\frac{2}{1+e^{-z}}\left(1-\frac{2}{1+e^{-z}}\right)\cdot x$$
Since, $b_1 \to b_1 - \lambda\frac{dL}{db_1}$, and $b_0 = 0, b_1 = -1, (x, y) = (1, 1)$.
Therfore, plug in the values, while setting $\lambda = 0.1$ for this gradient descent method, we get[2]:
$$b_1 \leftarrow -1 - 0.1\cdot 0.0053 \approx -1 - 0.00053 \approx -1$$

4.

The second-order Taylor series approximation of $L(b)$ around $\hat{b}$ is given by:
$$L(b) \approx L(\hat{b}) + (b-\hat{b})\frac{dL}{db}\bigg|_{\hat{b}} + \frac{1}{2}(b-\hat{b})^2\frac{d^2L}{db^2}\bigg|_{\hat{b}}$$
Next, take the first derivative and set it to zero to find minimum:
$$\frac{d}{db}\left(L(\hat{b}) + (b-\hat{b})\frac{dL}{db}\bigg|_{\hat{b}} + \frac{1}{2}(b-\hat{b})^2\frac{d^2L}{db^2}\bigg|_{\hat{b}}\right) = 0$$
Simplify,while ignoring the third derivative, and solve for b:
$$(b-\hat{b})\frac{d^2L}{db^2}\bigg|_{\hat{b}} = -\frac{dL}{db}\bigg|_{\hat{b}}$$
Rewritten in $b_1 \leftarrow b_1 - \lambda\frac{dL}{db_1}$ form, we get:
$$b = \hat{b} - \lambda\frac{dL}{db}\bigg|_{\hat{b}}$$
where $\lambda = \frac{1}{\frac{d^2L}{db^2}\big|_{\hat{b}}} = 1/L''(b)$.

3 Statistical learning

1.

Given the assumption, we have:
$$h_S(x) = \begin{cases} 1 & \text{if } \exists i \in [m] \text{ s.t. } x = x_i \text{ and } y_i = 1 \\ 0 & \text{otherwise} \end{cases}$$
It is easy to see that if $y_i = 0$, for all $i \in [m]$, then $h_S(x) = 0$, for all $x \in X$. If $p_S(x) := -1$, for all $x \in X$, then it is correct.
Inductively, Given: $S = \{(x_i, y_i)\}_{i=1}^m$, define the multivariate polynomial
$$p_S(x) = -\prod_{i\in[m]y_i=1} ||(x-x_i)^2||.$$
Then, for every i such that $y_i = 1$, we have $p_S(x_i) = 0$, while for all $i \in [m]$, we have $p_S(x) < 0$.

[2]with python