Machine Learning Assignment 35

Maia Dimas

Problem 35-1

A: Compute the likelihood P({1,17,8,25,3}|k). Remember that the likelihood is just the probability of getting the result {1,17,8,25,3} under the assumption that the data was sampled from the distribution $\mathbf{p}_k$(x). Your answer should be a piecewise function expressed in terms of k.

$$P(\{1,17,8,25,3\}|k) = \begin{cases} \frac{1}{k^5} & k \geq 25 \\ 0 & k < 25 \end{cases}$$

B: Compute the posterior distribution by normalizing the likelihood.

$$\sum_{k=1}^{\infty} c \cdot P(\{1,17,8,25,3\} \mid k) = c \cdot \sum_{k=25}^{\infty} \frac{1}{k^5}$$

$$c \cdot \sum_{k=25}^{\infty} \frac{1}{k^5} = 1$$

$$c \approx 1.4432 \cdot 10^6$$

$$P(k|\{1,17,8,25,3\}) = \begin{cases} \frac{1.4432 \cdot 10^6}{k^5} & k \geq 25 \\ 0 & k < 25 \end{cases}$$

C: What is the most probable value of k? You can tell this just by looking at the distribution p(k), but make sure to justify your answer with an explanation.

The most probable value of k is 25. This is just logic as 25 is the lowest possible number that can be in the denominator thus making the number as large as possible.

D: The largest number in the dataset is 25. What is the probability that 25 is actually the upper bound chosen by your friend?

Around 15%

$$\begin{aligned}
P(k = 25) &= P(25|\{1, 17, 8, 25, 3\}) \\
&= \frac{1.4432 \cdot 10^6}{25^5} \\
&= 0.14778368
\end{aligned}$$

E: What is the probability that the upper bound is less than or equal to 30?

Around a 58% chance

$$\begin{aligned}
P(k \leq 30) &= P(25 \leq k \leq 30) \\
&= \sum_{k=25}^{30} \frac{1.4432 \cdot 10^6}{k^5} \\
&= 0.583439
\end{aligned}$$

F: Fill in the blank: you can be 95% sure that the upper bound is less than ___.

53

Problem 35-2

A: Find the value of k such that p(x,y) is a valid probability distribution. Your answer should be in terms of a,b,c,d.

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \int_{-\infty}^{c}\left(\int_{-\infty}^{\infty} 0\,\mathrm{d}x\right)\,\mathrm{d}y + \int_{c}^{d}\left(\int_{-\infty}^{\infty} p(x,y)\,\mathrm{d}x\right)\,\mathrm{d}y + \int_{d}^{\infty}\left(\int_{-\infty}^{\infty} 0\,\mathrm{d}x\right)\,\mathrm{d}y \\
&= \int_{c}^{d}\left(\int_{-\infty}^{a} 0\,\mathrm{d}x + \int_{a}^{b} p(x,y)\,\mathrm{d}x + \int_{b}^{\infty} 0\,\mathrm{d}x\right)\,\mathrm{d}y \\
&= \int_{c}^{d}\int_{a}^{b} k\,\mathrm{d}x\,\mathrm{d}y \\
&= k\cdot(b-a)\cdot(d-c) \quad \text{(Simple integration)} \\
k\cdot(b-a)\cdot(d-c) &= 1 \\
k &= \frac{1}{(b-a)\cdot(d-c)}
\end{aligned}$$

B: Given that (X,Y) $\sim$ p, compute E[X] and E[Y].

$$\begin{aligned}
\mathrm{E}[X] &= \int_{c}^{d}\int_{a}^{b} x\cdot p(x,y)\,\mathrm{d}x\,\mathrm{d}y, \\
&= \left(\frac{1}{(b-a)\cdot(d-c)}\right)\cdot\int_{c}^{d}\int_{a}^{b} x\,\mathrm{d}x\,\mathrm{d}y, \\
&= \frac{(b^2-a^2)\cdot(d-c)}{2\cdot(b-a)\cdot(d-c)} \quad \text{(Simple Integration)} \\
&= \frac{(b-a)\cdot(b+a)}{2\cdot(b-a)} \\
&= \frac{b+a}{2}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[Y] &= \int_{c}^{d}\int_{a}^{b} y\cdot p(x,y)\,\mathrm{d}x\,\mathrm{d}y. \\
&= \left(\frac{1}{(b-a)\cdot(d-c)}\right)\cdot\int_{c}^{d}\int_{a}^{b} y\,\mathrm{d}x\,\mathrm{d}y, \\
&= \frac{(b-a)\cdot(d^2-c^2)}{2\cdot(b-a)\cdot(d-c)} \quad \text{(Simple Integration)} \\
&= \frac{(d-c)\cdot(d+c)}{2\cdot(d-c)} \\
&= \frac{d+c}{2}
\end{aligned}$$

C: Geometrically, $[a,b]\times[c,d]$ represents a rectangle bounded by $x = a$, $x = b$, $y = c$, and $y = d$. What is the geometric interpretation of the point $(E[X],E[Y])$ in this rectangle?

The center