Machine Learning HW1

Omar Ryyan – 322283755

Mahmoud Abadi – 206773756

Question 1:

Section A:

Similar to the example from tirgol, we compute $P(Sharp)$ and $P(Diminished)$ first by using total probability:

$$P(Sharp) = \sum_{class} P(Sharp|Class) * P(Class)$$

$$= P(Sharp|Good) * P(Good) + P(Sharp|Fair) * P(Fair) + P(Sharp|Bad) * P(Bad)$$
$$= 0.9 * 0.3 + 0.6 * 0.5 + 0.3 * 0.2 = 0.63$$

Hence $P(Diminished) = 1 - P(Sharp) = 1 - 0.63 = 0.37$

Now we calculate the posteriors probabilities given that the TV showed sharp picture:

$$P(Good|Sharp) = \frac{P(Sharp|Good)P(Good)}{P(Sharp)} = \frac{0.9 * 0.3}{0.63} = \frac{0.27}{0.63}$$

$$P(Fair|Sharp) = \frac{P(Sharp|Fair)P(Fair)}{P(Sharp)} = \frac{0.6 * 0.5}{0.63} = \frac{0.3}{0.63}$$

$$P(Bad|Sharp) = \frac{P(Sharp|Bad)P(Bad)}{P(Sharp)} = \frac{0.3 * 0.2}{0.63} = \frac{0.06}{0.63}$$

Now we calculate the posteriors probabilities given that the TV showed diminished picture:

$$P(Good|Diminished) = \frac{P(Diminished|Good)P(Good)}{P(Diminished)} = \frac{0.1 * 0.3}{0.37} = \frac{0.03}{0.37}$$

$$P(Fair|Diminished) = \frac{P(Diminished|Fair)P(Fair)}{P(Diminished)} = \frac{0.4 * 0.5}{0.37} = \frac{0.2}{0.37}$$

$$P(Bad|Diminished) = \frac{P(Diminished|Bad)P(Bad)}{P(Diminished)} = \frac{0.7 * 0.2}{0.37} = \frac{0.14}{0.37}$$

Now we calculate the conditional risk for buying or not depending on the picture on the TV.

Sharp picture:

$$R(Buy|\text{Sharp})$$

$$= \lambda(Buy|Good) * P(Good|Sharp) + \lambda(Buy|Fair) * P(Fair|Sharp) + \lambda(Buy|Bad) * P(Bad|Sharp) = 0 * \frac{0.27}{0.63} + 4 * \frac{0.3}{0.63} + 20 * \frac{0.06}{0.63} = 3.81$$

$$R(Dont\ Buy|\text{Sharp})$$

$$= \lambda(Dont\ Buy|Good) * P(Good|Sharp) + \lambda(Dont\ Buy|Fair) * P(Fair|Sharp) + \lambda(Dont\ Buy|Bad) * P(Bad|Sharp) = 10 * \frac{0.27}{0.63} + 6 * \frac{0.3}{0.63} + 0 * \frac{0.06}{0.63} = 7.14$$

Thus, in the case of a sharp picture, the optimal decision is to buy the TV.

Diminished picture:

$$R(Buy|Diminished)$$

$$= \lambda(Buy|Good) * P(Good|Diminished) + \lambda(Buy|Fair) * P(Fair|Diminished) + \lambda(Buy|Bad) * P(Bad|SharpDiminished) = 0 * \frac{0.03}{0.37} + 4 * \frac{0.2}{0.37} + 20 * \frac{0.14}{0.37} = 9.73$$

$$R(Dont\ Buy|Diminished)$$

$$= \lambda(Dont\ Buy|Good) * P(Good|Diminished) + \lambda(Dont\ Buy|Fair) * P(Fair|Diminished) + \lambda(Dont\ Buy|Bad) * P(Bad|Diminished) = 10 * \frac{0.03}{0.37} + 6 * \frac{0.2}{0.37} + 0 * \frac{0.14}{0.37} = 4.05$$

Thus, in the case of a diminished picture, the optimal decision is to not buy the TV.

Section B:

Since we use Bayesian decision rule and we define the 0-1 loss function such that $\lambda(a_i|c_j) = \begin{cases} 1 & if \quad i = j \\ 0 & if \quad i \neq j \end{cases}$, we plug the 0-1 loss into the conditional risk formula:

$$R(\alpha_i|x) = \sum_j \lambda(a_i|c_j) * P(c_j|x) = \sum_{j=i} \lambda(a_i|c_j) * P(c_j|x) + \sum_{j\neq i} \lambda(a_i|c_j) * P(c_j|x)$$

$$= \sum\nolimits_{j=i} 0 * P(c_j|x) + \sum\nolimits_{j\neq i} 1 * P(c_j|x) = \sum\nolimits_{j\neq i} 1 * P(c_j|x)$$

Since the sum of all posterior probabilities is 1, we get:

$$\sum_{j\neq i} 1 * P(c_j|x) = 1 - P(c_i|x)$$

Thus:

$$R(\alpha_i|x) = 1 - P(c_i|x)$$

And minimizing $R(\alpha_i|x)$ is the same as maximising $P(c_i|x)$, thus under the 0-1 loss function the Bayesian decision rule is equivalent to the MAP classification.

Question 2:

PS: The class c determines r_c, we can type $L(r_c) = \prod_{i=1}^{n} P(x_i|r_c)$ but we feel it's correct to stay with $L(c) = \prod_{i=1}^{n} P(x_i|c)$.

$C - number\ of\ classes$

We first type the likelihood function:

$$L(c) = \prod_{i=1}^{n} P(x_i|c) = \prod_{i=1}^{n} \frac{2x_i}{r_c^2} * I_{0\leq x_i \leq r_c}$$

Meaning that $P(x_i|c) = 0$ if any $x_i > r_c$, hence if any x_i is bigger than r_c then $\prod_{i=1}^{n} P(x_i|c) = 0$, thus $r_c \geq \max_i(x_i)$. Now we look at the likelihood again:

$$L(c) = \prod_{i=1}^{n} \frac{2x_i}{r_c^2} * I_{\max x_i \leq r_c} = \left(\prod_{i=1}^{n} 2x_i\right) * r_c^{-2n} * I_{\max x_i \leq r_c}$$

We notice that $\prod_{i=1}^{n} 2x_i$ is constant, r_c^{-2n} decreases as r_c increases and $I_{\max x_i \leq r_c}$ is zero unless $\max_i(x_i) \leq r_c$ so to maximize the likelihood we need the minimum r_c which is $r_c = \max_i(x_i; x_i \in c)$ since we can't take higher r_c because r_c^{-2n} decreases and we can't take lower because likelihood becomes zero.

We notate $r_c = \max_i(x_i; x_i \in c) = \hat{r}_c$, then:

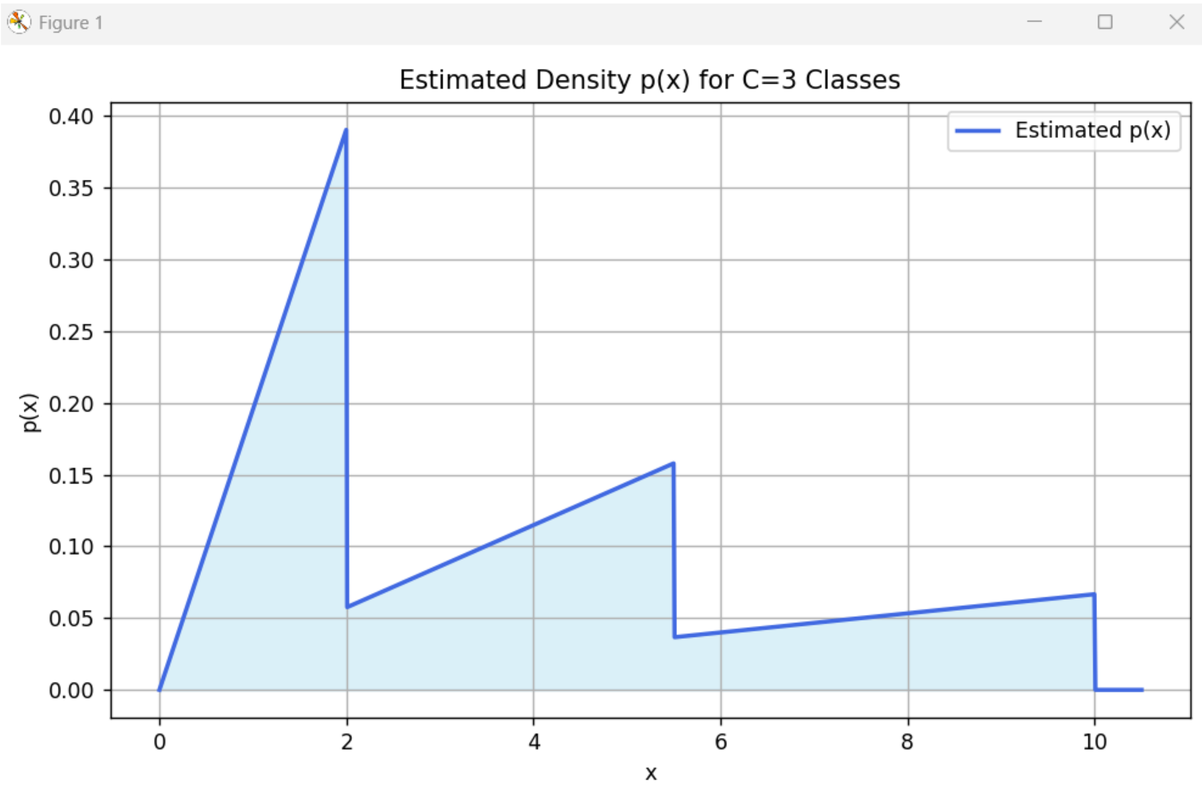
$$P(x) = \sum_{c=1}^{C} P(x|c) * P(c) = \sum_{c=1}^{C} \frac{2x}{\hat{r}_c^2} * I_{0 \le x \le \hat{r}_c} * P(c)$$

We assume that all classes are equally likely, thus $P(c) = \frac{1}{C}$, which makes $P(x)$:

$$P(x) = \sum_{c=1}^{C} \frac{2x}{\hat{r}_c^2} * I_{0 \le x \le \hat{r}_c} * \frac{1}{C} = \sum_{c=1}^{C} \frac{2x}{C * \hat{r}_c^2} * I_{0 \le x \le \hat{r}_c}$$

To draw $p(x)$ we will use the following fixed samples:

$C = 3$,$\hat{r}_1 = 2$, $\hat{r}_2 = 10$ and $\hat{r}_3 = 5.5$ with the functions being $f_i(x) = \frac{2x_i}{3 * \hat{r}_i^2} * I_{0 \le x \le \hat{r}_i}$ for $i \in \{1,2,3\}$ and $p(x) = f_1(x) + f_2(x) + f_2(x)$. Using python to draw the function:



Now we assume that $r_c \sim U(c,5)$, we find $P(x)$ using Bayesian density Estimation:

$$P(x) = \int_{\max(c,\max x_i,x)}^{5} P(x|r_c)P(r_c|D)dr_c$$

r_c must be more than $\max(c, \max x_i\,, x)$ otherwise the integrand is 0.

We know that $P(x|r_c) = \frac{2x}{r_c^2} * I_{0\le x\le \hat{r}_c}$.

From Bayesian Theorem:

$$P(r_c|D) = \frac{P(D|r_c) * P(r_c)}{P(D)}$$

We calculated the likelihood above which is $P(D|r_c)$:

$$P(D|r_c) = \prod_{i=1}^{n} \frac{2x_i}{r_c^2} * I_{\max x_i \le r_c}$$

Since $r_c \sim U(c,5)$:

$$P(r_c) = \begin{cases} \dfrac{1}{5-c} & if\ r_c \in [c,5] \\ 0 & otherwise \end{cases}$$

To find $P(D)$ we do the following:

$$P(D) = \int_{\max(c,\max x_i)}^{5} P(D|r_c)P(r_c)dr_c$$

Which is a constant, and since we don't have exact values of c and x_i we leave it as Z.

We substitute back into the equation of $P(r_c|D)$:

$$P(r_c|D) = \frac{P(D|r_c) * P(r_c)}{P(D)} = \frac{\prod_{i=1}^{n} \frac{2x_i}{r_c^2} * I_{\max x_i \le r_c} * \frac{1}{5-c}}{Z}$$

We type it in a cleaner way:

$$P(r_c|D) = \begin{cases} \dfrac{\prod_{i=1}^{n} \frac{2x_i}{r_c^2}}{Z(5-c)} & if\ \max(c, \max x_i) \le r_c \le 5 \\ 0 & otherwise \end{cases}$$

We substitute back into the equation of $P(x)$:

$$P(x) = \int_{\max(c,\max x_i,x)}^{5} P(x|r_c)P(r_c|D)dr_c = \int_{\max(c,\max x_i,x)}^{5} \frac{2x}{r_c^2} * \frac{\prod_{i=1}^{n} \frac{2x_i}{r_c^2}}{Z(5-c)} dr_c$$

If $x \leq \max x_i$ for each i, then $P(x) = \alpha$ (constant) since it won't depend on x.

Same for if $x \leq c$.

If $x > \max x_i$ for each i and $x > c$, then:

$$P(x) = \int_{x}^{5} \frac{2x}{r_c^2} * \frac{\prod_{i=1}^{n} \frac{2x_i}{r_c^2}}{Z(5-c)} dr_c = \frac{2x * \prod_{i=1}^{n} 2x_i}{Z(5-c)} \int_{x}^{5} \frac{1}{r_c^2 * \prod_{i=1}^{n} r_c^2} dr_c$$

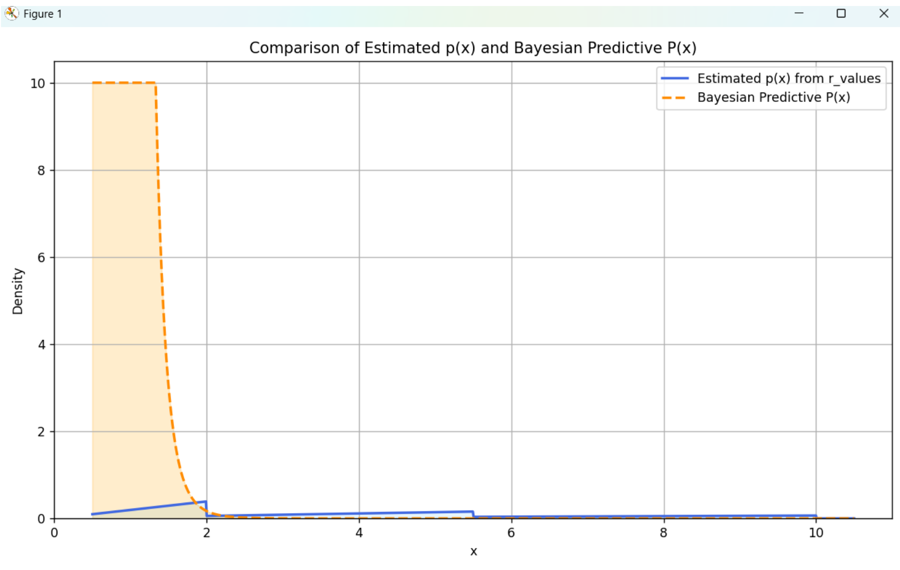
Lets compute the integral alone:

$$\int_{x}^{5} \frac{1}{r_c^2 * \prod_{i=1}^{n} r_c^2} dr_c = \int_{x}^{5} \frac{1}{r_c^{2(n+1)}} dr_c = \int_{x}^{5} r_c^{-2n-2} dr_c = \left[\frac{r_c^{-2n-1}}{-2n-1}\right]_x^5$$

$$= \frac{1}{-2n-1} * (5^{-2n-1} - x^{-2n-1}) = \frac{1}{2n+1} * (x^{-2n-1} - 5^{-2n-1})$$

Then $P(x) = \frac{2x * \prod_{i=1}^{n} 2x_i}{Z(5-c)*(2n+1)} * (x^{-2n-1} - 5^{-2n-1})$.

Using a python code with fixed examples, we plotted p(x) from this method together with the previous method:



We would say the second method (Bayesian Estimation) is better because the resulting $P(x)$ is smoother and it doesn't blindly trust the data and keeps uncertainty about r_c in mind.