Lab 2: Exercises

ML701: Machine Learning
MBZUAI

September 5, 2024

Problem 1. Suppose $\mathbb{P}[A], \mathbb{P}[A'], \mathbb{P}[B|A]$, and $\mathbb{P}[B|A']$ are known, where A' is a complement of A, i.e., $A' = \mathcal{X} \setminus A$. Find an expression for $\mathbb{P}[A|B]$ in terms of these four probabilities.

Solution. Using Bayes' rule with the fact that $\mathbb{P}[A' \cup A] = 1$, i.e., it is the certain event, we have

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cup B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[A]\,\mathbb{P}[B|A]}{\mathbb{P}[A]\,\mathbb{P}[B|A] + \mathbb{P}[A']\,\mathbb{P}[B|A']}.$$

Problem 2. Assume the probability of having tuberculosis (TB) is 0.0005, and a test for TB is 99% accurate. What is the probability one has TB if one tests positive for the disease?

Solution. We apply the solution to Problem 1 with A correspond to having tubercolosis, and B refers to tubercolosis test being positive. Then, we can rewrite the problem formulation as

$$\mathbb{P}[A] = 0.005, \quad \mathbb{P}\left[A'\right] = 0.995, \quad \mathbb{P}[B|A] = 0.99, \quad \text{and} \quad \mathbb{P}\left[B|A'\right] = 0.01.$$

Therefore, plugging this to the derived formula

$$\begin{aligned}\mathbb{P}[\text{Having Tuberculosis}|\text{Positive Test}] = \mathbb{P}[A|B] &= \frac{\mathbb{P}[A]\,\mathbb{P}[B|A]}{\mathbb{P}[A]\,\mathbb{P}[B|A] + \mathbb{P}[A']\,\mathbb{P}[B|A']} \\ &= \frac{0.005 * 0.99}{0.005 * 0.99 + 0.995 * 0.01} \approx 0.0472.\end{aligned}$$

Problem 3. [**Coupon Collector's Problem**]
Imagine you're collecting trading cards from cereal boxes. There are n distinct cards to collect, and each cereal box contains one card, randomly chosen with equal probability. The question is: on average, how many cereal boxes do you need to buy in order to collect all n distinct cards?
Hint: Try to use linearity of expectation.

Solution. For card i, let X_i be the number of boxes you have to buy after collecting the $i-1$ distinct card until you get the i^{th} distinct card. The goal is to find $\mathbb{E}[X]$ where $X = X_1 + X_2 + ... + X_n$.
Notice:

1. $\mathbb{E}[X_1] = 1$ because you are guaranteed to get a new card in the first box.
2. After obtaining the first card, the probability that the next box contains a new card is $\frac{n-1}{n}$. Therefore, $\mathbb{E}[X_2] = \frac{n}{n-1}$.
3. After obtaining the first two cards, the probability that the next box contains a new card is $\frac{n-2}{n}$. Therefore, $\mathbb{E}[X_3] = \frac{n}{n-2}$.

4. And so on...

Therefore, the expected number of boxes to get all n distinct cards is:

$$\begin{aligned}\mathbb{E}\left[X\right] &= \mathbb{E}\left[X_1 + X_2 + ... + X_n\right] \\ &= \mathbb{E}\left[X_1\right] + \mathbb{E}\left[X_2\right] + ... + \mathbb{E}\left[X_n\right] \\ &= 1 + \frac{n}{n-1} + \frac{n}{n-2} + ... + n \\ &= n\left(1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{n}\right).\end{aligned}$$

The sum in the parentheses is the n^{th} Harmonic number, which is approximately $\log(n) + \gamma$ where γ is the Euler-Mascheroni constant.

Problem 4. [**Decision Rule Based on Posterior Probabilities and Joint Densities**] Consider a binary prediction problem where the outcome Y can take values in the set {0,1}. Suppose we're making decisions based on the ratio:

$$\frac{\mathbb{P}\left[Y=1|X=x\right]}{\mathbb{P}\left[Y=0|X=x\right]}$$

Prove that any decision rule using the above ratio is equivalent to considering the ratio of the joint densities:

$$\frac{p(x, Y=1)}{p(x, Y=0)}$$

Solution. Using Bayes' rule, we have

$$\frac{\mathbb{P}\left[Y=1|X=x\right]}{\mathbb{P}\left[Y=0|X=x\right]} = \frac{\frac{p(x,Y=1)}{p(x)}}{\frac{p(x,Y=0)}{p(x)}} = \frac{p(x, Y=1)}{p(x, Y=0)}.$$

Problem 5. [**Optimal Binary Prediction with Symmetric Loss and MAP Rule**] Consider a binary prediction problem, where the outcome Y can take values in the set {0,1}. Suppose we have a symmetric loss function, i.e.,

$$\text{loss}(y, y') = \text{loss}(y', y)$$

for all y, y' in $\{0, 1\}$. Furthermore, assume $\text{loss}(1,1) = \text{loss}(0,0) = 0$.

1. Show that for every such symmetric loss, the optimal binary predictor is equivalent to the Maximum A Posteriori (MAP) rule.
2. Furthermore, if $\mathbb{P}\left[Y=1\right] = \mathbb{P}\left[Y=0\right] = \frac{1}{2}$, demonstrate that the likelihood estimator yields the same result as the MAP rule.

Solution. For the first part, using Lemma 16, we have

$$\begin{aligned}\hat{Y}^{\star}(x) &= \mathbb{1}\left\{\mathbb{P}\left[Y=1 \mid X=x\right] \geq \frac{loss(1,0) - loss(0,0)}{loss(0,1) - loss(1,1)} \cdot \mathbb{P}\left[Y=0 \mid X=x\right]\right\} \\ &\overset{\text{loss}(1,1)=\text{loss}(0,0)=0}{=} \mathbb{1}\left\{\mathbb{P}\left[Y=1 \mid X=x\right] \geq \frac{loss(1,0)}{loss(0,1)} \cdot \mathbb{P}\left[Y=0 \mid X=x\right]\right\} \\ &\overset{\text{loss}(1,0)=\text{loss}(0,1)}{=} \mathbb{1}\left\{\mathbb{P}\left[Y=1 \mid X=x\right] \geq \cdot \mathbb{P}\left[Y=0 \mid X=x\right]\right\} \\ &= \arg\max_{y\in\{0,1\}} \mathbb{P}\left[Y=y \mid X=x\right].\end{aligned}$$

For the second part,

$$
\begin{aligned}
\hat{Y}^{\star}(x) &= \arg\max_{y\in\{0,1\}} \mathbb{P}\left[Y = y \mid X = x\right] \\
&= \arg\max_{y\in\{0,1\}} \frac{p(x, Y = y)}{p(x)} \\
&\overset{p(x)\perp y \text{ appears in both}}{=} \arg\max_{y\in\{0,1\}} p(x, Y = y) \\
&\overset{\text{Bayes' rule}}{=} \arg\max_{y\in\{0,1\}} \mathbb{P}\left[Y = y\right] p(x|Y = y) \\
&\overset{\mathbb{P}[Y=y]=1/2,\ \forall y\in\{1,2\}}{=} \arg\max_{y\in\{0,1\}} p(x|Y = y).
\end{aligned}
$$