CSE 250B: Section 1 - Sharad Vikram

1. Let X and Y be binary random variables ($X, Y \in \{0, 1\}$). Given the following, calculate $\mu(x)$ and $h(x)$:

$$\begin{aligned} P(Y = 0) &= 0.3 \\ P(Y = 1) &= 0.7 \\ P(X = 1|Y = 0) &= 0.4 \\ P(X = 1|Y = 1) &= 0.9 \end{aligned}$$

Solution: We can calculate the joint distribution of X and Y by calculating $P(Y = y, X = x) = P(X = x|Y = y)P(y = y)$ for all values of X and Y.

$$\begin{aligned} P(Y = 0, X = 0) &= P(X = 0|Y = 0)P(Y = 0) \\ &= (1 - 0.4) \times 0.3 = 0.18 \\ P(Y = 1, X = 0) &= P(X = 0|Y = 1)P(Y = 1) \\ &= (1 - 0.9) \times 0.7 = 0.07 \\ P(Y = 0, X = 1) &= P(X = 1|Y = 0)P(Y = 0) \\ &= 0.4 \times 0.3 = 0.12 \\ P(Y = 1, X = 1) &= P(X = 1|Y = 1)P(Y = 1) \\ &= 0.9 \times 0.7 = 0.63 \end{aligned}$$

Using the joint distribution, we can marginalize and divide to calculate $\mu(x) = P(Y = 1|X = x)$. Note that we need to calculate $\mu(x)$ separately for $X = 0$ and $X = 1$.

$$\begin{aligned} P(Y = 1|X = 0) &= \frac{P(Y = 1, X = 0)}{P(X = 0)} \\ &= \frac{0.07}{0.07 + 0.18} \\ &= 0.28 \\ P(Y = 1|X = 1) &= \frac{P(Y = 1, X = 1)}{P(X = 1)} \\ &= \frac{0.63}{0.63 + 0.12} \\ &= 0.84 \end{aligned}$$

We get a final solution of:

$$\mu(x) = \begin{cases} 0.28 & \text{if } x = 0 \\ 0.84 & \text{if } x = 1 \end{cases}$$

To obtain $h(x)$, we decide $y = 1$ if $\mu(x) > 0.5$ and $y = 0$ otherwise.

$$h(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{if } x = 1 \end{cases}$$

2. You are given the following for $x \in [0, 1]$:

$$\eta(x) = \begin{cases} 0.9 & x \leq 0.5 \\ 0.4 & x > 0.5 \end{cases}$$
$$\mu(x) = 2x$$

a) What is $h^*(x)$?

Solution: We can just threshold $\mu(x)$ at 0.5 to find $h(x)$.

$$h(x) = \begin{cases} 1 & \text{if } x \leq 0.5 \\ 0 & \text{if } x > 0.5 \end{cases}$$

b) What is R^*?

Solution: The Bayes risk of our classifier $R^* = R(h^*)$ is the expected *conditional risk*, or $\mathbb{E}_x[R(h|x)]$. The conditional risk is the risk associated with a particular data point with value x. For the standard classification scenario (zero-one loss), where we aim to minimize the amount of errors we make, the conditional risk formula is $R(h|x) = \min(\eta(x), 1 - \eta(x))$.
We can thus calculate R^* with the following integral:

$$
\begin{aligned}
R^* &= \mathbb{E}_x[\min(\eta(x), 1-\eta(x))] \\
&= \int_0^1 p(x)\min(\eta(x), 1-\eta(x))dx \\
&= \int_0^{0.5} 2x \times 0.1dx + \int_{0.5}^1 2x \times 0.4dx \\
&= 0.025 + 0.3 \\
&= 0.325
\end{aligned}
$$

c) What are h^* and R^* if $\mu(x) = 1$?

Solution: h^* remains the same since it depends purely on $\eta(x)$.

$$
\begin{aligned}
R^* &= \mathbb{E}_x[\min(\eta(x), 1-\eta(x))] \\
&= \int_0^1 p(x)\min(\eta(x), 1-\eta(x))dx \\
&= \int_0^{0.5} 0.1dx + \int_{0.5}^1 0.4dx \\
&= 0.25
\end{aligned}
$$

3. Prove that the following function is a distance metric:

$$d(x, y) = \max_i(|x_i - y_i|)$$

Solution: There are 4 conditions for a function to be a distance metric:

(a) $d(x, y) \geq 0$

This is true, since this function takes the absolute value of the elements of a vector and picks the largest one, always resulting in a non-negative number.

(b) $d(x, y) = 0$ if and only if $x = y$.

Right direction: Proof by contradiction. If $x \neq y$, there exists a non-zero vector difference between the two vectors. The absolute value of the largest element of the vector is therefore a positive number, and not zero.

Left direction: Proof by contradiction. If $d(x, y) \neq 0$, there exists a pair of elements in x and y, x_j and y_j such that $x_j - y_j \neq 0$. This implies that $x \neq y$.

(c) $d(x, y) = d(y, x)$

This is true because $|x_i - y_i| = |y_i - x_i|$.

(d) $d(x, z) \leq d(x, y) + d(y, z)$

This function is known as l_∞ distance, based on the l_∞ norm.

Lemma: $||x + y||_\infty \leq ||x||_\infty + ||y||_\infty$ (prove this for yourself)

Assuming the above lemma:

$$\begin{aligned} d(x, y) &= ||x - z||_\infty \\ &= ||x - z + y - y||_\infty \\ &= ||(x - y) + (y - z)||_\infty \\ &\leq ||(x - y)||_\infty + ||(y - z)||_\infty \\ &= d(x, y) + d(y, z) \end{aligned}$$