Assignment 2

Muhammed Saeed and Ali Salaheldin Ali Ahmed

November 2022

Problem 2

2.1.a.

We start by expressing the posteriors:

$$\begin{aligned} P(Y=1|\,X) &= P(X|\,Y=1)P(Y=1) \\ P(Y=0|\,X) &= P(X|\,Y=0)P(Y=0) \\ &= cP(X|\,Y=0)P(Y=1) \end{aligned}$$

Now, let us consider the case of $\|X\| > r$:

$$\begin{aligned} &P(Y=1|\,X) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{\|X\|^2}{2\sigma^2}\right)P(Y=1) \\ &P(Y=0|\,X) = 0 \end{aligned}$$

Since we can safely assume that $P(Y=1) > 0$ (otherwise $P(Y=1) = P(Y=0) = 0$), the optimal prediction in this case is $Y=1$.

Next, for $\|X\| \le r$:

$$\begin{aligned} &P(Y=1|\,X) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{\|X\|^2}{2\sigma^2}\right)P(Y=1) \\ &P(Y=0|\,X) = c\frac{1}{\pi r^2}P(Y=1) \end{aligned}$$

We predict that $Y = 1$ if $P(Y = 1|\, X) \geq P(Y = 0|\, X)$, that is, if:

$$\begin{aligned}
\exp\left(-\frac{\|X\|^2}{2\sigma^2}\right) &\geq 2c\frac{\sigma^2}{r^2} \\
-\frac{\|X\|^2}{2\sigma^2} &\geq \ln(c) + \ln(\sigma^2) - \ln(\frac{r^2}{2}) \\
\frac{\|X\|^2}{2\sigma^2} &< \ln(\frac{r^2}{2}) - \ln(c) - \ln(\sigma^2) \\
\|X\|^2 &< 2\sigma^2\left(\ln(\frac{r^2}{2}) - \ln(c) - \ln(\sigma^2)\right)
\end{aligned}$$

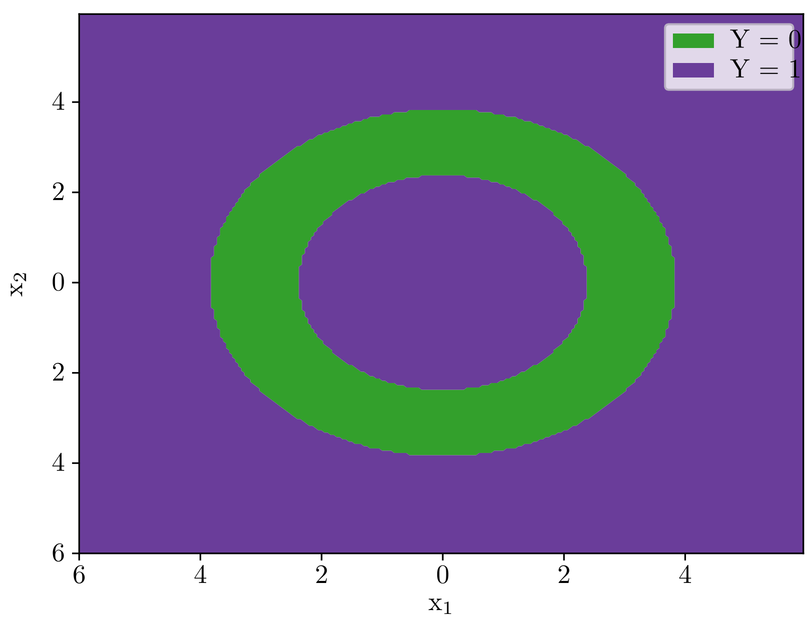
otherwise, we predict $Y = 0$.
Overall, the Bayes optimal classifier can be expressed as:

$$\hat{y} = \begin{cases} 0 & (\|X\| \leq r) \wedge (\|X\|^2 \geq 2\sigma^2\left(\ln(\frac{r^2}{2}) - \ln(c) - \ln(\sigma^2)\right)) \\ 1 & \text{otherwise} \end{cases}$$

2.1.b.

Substituting the values into the conditions from the previous formula:

$$\hat{y} = \begin{cases} 0 & \sqrt{\frac{14}{3}} \leq \|X\| \leq e\sqrt{2} \\ 1 & \text{otherwise} \end{cases}$$



The decision boundary consists of two circular curves, where the inner donut shaped region corresponds to $Y = 0$ and the outer area $Y = 1$.

The first part of the inequality, which controls the radius of inner circle of the decision boundary, can be formulated in terms of only c as:

$$\|X\| \geq \sqrt{2}\sqrt{2 - \ln c}$$

We can see that increasing c would decrease the inner radius of the boundary, which would mean more X values would be classified as $Y = 0$. We think this makes sense, since increasing c means increasing $P(Y = 0)$

2.2.

- Conditional Independence The assumption is not always accurate. The feature typically exhibits some sort of dependency.
- Zero probability issue: We may have zero class probabilities if we come across terms in the test data for a given class that are absent from the training data.This can be handled to some extent by adding smoothing factor