1. Logistic posterior with exponential class conditionals

   We have seen in class that Gaussian class conditionals lead to a logistic posterior that is quadratic in X. Now, suppose the class conditionals are exponentially distributed with parameteres λ_i, i.e.

$$\begin{aligned} X|Y = i &\sim \lambda_i \exp(-\lambda_i x), \quad \text{where } i \in \{0, 1\} \\ Y &\sim \text{Bernoulli}(\pi) \end{aligned}$$

   Show that the posterior distribution of the class label given X is also a logistic function, however with a linear argument in X. Assuming 0-1 loss, what will the decision boundary look like (i.e., describe what the posterior probability plot looks like)?

Solution:

We are solving for $P(Y = 1|x)$. By Bayes Rule, we have

$$\begin{aligned} P(Y = 1|x) &= \frac{P(x|Y = 1)\, P(Y = 1)}{P(x|Y = 1)\, P(Y = 1) + P(x|Y = 0)\, P(Y = 0)} \\ &= \frac{1}{1 + \frac{P(Y=0)P(x|Y=0)}{P(Y=1)P(x|Y=1)}} \\ &= \frac{1}{1 + \frac{\lambda_0}{\lambda_1}\frac{1-\pi}{\pi}\exp(-\lambda_0 x + \lambda_1 x)} \end{aligned}$$

Looking at the bottom right equation, we have

$$\frac{\lambda_0}{\lambda_1}\frac{1-\pi}{\pi}\exp(-\lambda_0 x + \lambda_1 x) = \exp\left(-(\lambda_0 - \lambda_1)x + \log\left(\frac{\lambda_0}{\lambda_1}\frac{1-\pi}{\pi}\right)\right)$$

Now we see that we have a logistic function $\frac{1}{1+\exp(-h(x))}$, where $h(x) = ax + b$ is linear (affine) in x. Since we are assuming 0-1 loss, we use the optimal classifier $f^*(x) = 1$ when $P(Y = 1|x) > P(Y = 0|x)$. Thus, the decision boundary can be found when $P(Y = 1|x) = P(Y = 0|x) = \frac{1}{2}$. This happens when $h(x) = 0$. Solving for x gives

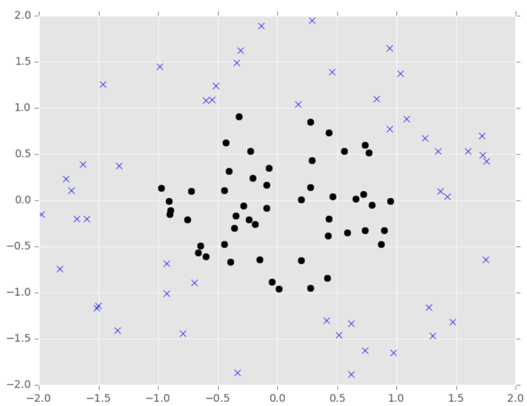
$$\bar{x} = \frac{\log \frac{\lambda_0}{\lambda_1}\frac{1-\pi}{\pi}}{\lambda_0 - \lambda_1}.$$

If we assume $\lambda_0 > \lambda_1$, then the optimal classifier is

$$f^*(x) = \begin{cases} 1 & \text{if } x > \bar{x} \\ 0 & o.w. \end{cases}$$

2. Circular Distributions

   Consider the following dataset



where each point $x_n = (x_{1,n}, x_{2,n})$ is sampled *iid* and uniformly at random from two equiprobable (each equally likely) classes, a disk of radius 1 ($y_n = 1$) and a ring from 1 to 2 ($y_n = -1$).

Qualitatively estimate and sketch the following quantities

- The class conditional densities $f_{X|Y=c_i}$
- The density of X
- The posterior probabilities $P(Y = c_i|X)$
- The Bayes Classifier
- The Bayes Risk

Solution:

- The class conditionals:

$$f_{X|Y=1} = \begin{cases} \frac{1}{\pi} & \|X\|_2 \leq 1 \\ 0 & o.w. \end{cases}$$

$$f_{X|Y=-1} = \begin{cases} \frac{1}{3\pi} & 1 \leq \|X\|_2 \leq 2 \\ 0 & o.w. \end{cases}$$

- The density of X:

$$f_X = \frac{1}{2}\left(f_{X|Y=1} + f_{X|Y=-1}\right)$$

- The posterior probabilities $P(Y|X)$:

$$P(Y = 1 \mid X) = \begin{cases} 1 & \|X\|_2 \leq 1 \\ 0 & o.w. \end{cases}$$

$$P(Y = -1 \mid X) = \begin{cases} 1 & 1 \leq \|X\|_2 \leq 2 \\ 0 & o.w. \end{cases}$$

- Bayes Classifier: the circle $x_1^2 + x_2^2 = 1$
- The Bayes Risk = 0

3. Bayesian Decision Theory: Case Study - We're going fishing!

We want to design an automated fishing system that captures fish, classifies them, and sends them off to two different companies, Salmonites, Inc., and Seabass, Inc. For some reason we only ever catch salmon ($Y = 1$) and seabass ($Y = 2$). Salmonites, Inc. wants salmon, and Seabass, Inc. wants seabass. Given only the weights of the fish we catch, we want to figure out what type of fish it is using machine learning!

Let us assume that the weight of both seabass and salmon are both normally distributed (univariate Gaussian), given by the p.d.f.:

$$P(x|Y = i) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}$$

We are given this data:

Data for salmon: $\{3, 4, 5, 6, 7\}$

Data for seabass: $\{5, 6, 7, 8, 9, 7 + \sqrt{2}, 7 - \sqrt{2}\}$

When we classify seabass incorrectly, it gets sent to Salmonites, Inc. who won't pay us for the wrong fish and sells it themselves. When we classify salmon incorrectly, it gets sent to SeaBass, Inc., who is nice and returns our fish. This situation gives rise to this loss matrix:

Predicted:

Truth:		**salmon**	**seabass**
	salmon	0	1
	seabass	2	0

a) First, compute the sample mean $\hat{\mu}_i$ and variance $\hat{\sigma}_i^2$ for the univariate Gaussian in both the seabass and the salmon case. Also compute the empirical estimates of the priors $\hat{\pi}_i$.

Solution: Sample mean $\hat{\mu}$:

$$\hat{\mu} = \frac{1}{N}\sum_i X_i$$

Sample covariance:

$$\hat{\sigma}^2 = \frac{1}{N}\sum_i (X_i - \hat{\mu})^2$$

Plugging in numbers for seabass and salmon: $\hat{\mu}_1 = 5$, $\hat{\mu}_2 = 7$, $\hat{\sigma}_1^2 = 2$, $\hat{\sigma}_2^2 = 2$

Calculating the priors: $\hat{\pi}_1 = 5/12$, $\hat{\pi}_2 = 7/12$

What is significant about $\hat{\sigma_1}$ and $\hat{\sigma_2}$?

Solution: They're the exact same, so a decision boundary between the two Gaussians characterized by them will be linear.

b) Next, find the decision rule when assuming a 0-1 loss function. Recall that a decision rule for the 0-1 loss function will minimize the probability of error.

Solution: Recall that assuming a 0-1 loss function results in choosing the class to minimize the probability of error, which means choosing according to this rule:

$$\text{If } \frac{p(Y=1|x)}{p(Y=2|x)} > 1, \text{ choose } 1$$

Because there is a linear decision boundary, we search for the value such that we classify everything to the right as seabass, and everything to the left as salmon. This boundary is the value of x such that $p(Y=1|x) = p(Y=2|x)$.

$$p(Y=1|x) = p(Y=2|x) \implies 5p(x|Y=1) = 7p(x|Y=2)$$

$$\frac{5}{\sqrt{2\pi}\sigma} exp\left(-\frac{1}{2}\frac{(x-5)^2}{\sigma^2}\right) = \frac{7}{\sqrt{2\pi}\sigma} exp\left(-\frac{1}{2}\frac{(x-7)^2}{\sigma^2}\right)$$

$$\ln(5) - \frac{1}{2\sigma^2}(x-5)^2 = \ln(7) - \frac{1}{2\sigma^2}(x-7)^2$$

$$4\ln\left(\frac{5}{7}\right) - x^2 + 10x - 25 = -x^2 + 14x - 49$$

$$4\ln\left(\frac{5}{7}\right) + 24 = 4x$$

$$x = \ln\left(\frac{5}{7}\right) + 6 \approx 5.66$$

The decision rule is: If $x > 5.66$, classify as Seabass! Otherwise classify as Salmon.

Note: Because we had the same variance for both class conditionals, the x^2 term canceled out. If that was not the case, then there would be 3 regions, and we would allocate 2 of them to one fish, 1 of them to the other, depending on the height of the posterior probabilities. A good exercise would be to try to draw this: two 1-D Gaussians with different variances.

c) Now, find the decision rule using the loss matrix above. Recall that a decision rule, in general, minimizes the risk, or expected loss.

Solution: In the general case, we want to make the decision that minimizes risk. Thus, the decision boundary is located at where the risk of making either decision is equal, or:

$$R(\hat{y} = 1|x) = R(\hat{y} = 2|x)$$

Recall that $R(\hat{y} = i|x) = \sum_{j=1}^{C} \lambda_{ij} P(Y = j|x)$.

$$\lambda_{11} P(Y = 1|x) + \lambda_{12} P(Y = 2|x) = \lambda_{21} P(Y = 1|x) + \lambda_{22} P(Y = 2|x)$$

$$2 \cdot P(Y = 2|x) = 1 \cdot P(Y = 1|x)$$

$$2 \cdot \frac{7}{12} \mathcal{N}(7, 2) = 1 \cdot \frac{5}{12} \mathcal{N}(5, 2)$$

Solving this like part b), we get that $x = 6 + \ln(\frac{5}{14}) \approx 4.97$. Thus, if the weight is greater than 4.97, we classify it as seabass and if not, we classify it as salmon.