Bayesian classification example

Dror Baron

This supplement works through an example on Bayesian classification.

Problem: Consider a classification problem, where the data is labeled as either red or blue. The probability of each color is half, i.e., Pr(red) = Pr(blue) = 0.5. The two classes correspond to pdfs

$$f_{\text{blue}} = 0.5\mathcal{N}(-1,1) + 0.5\mathcal{N}(2,1)$$

and

$$f_{\text{red}} = \mathcal{N}(0,1),$$

where $\mathcal{N}(\mu,\sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2, in our specific pdfs all variances are the same while the means differ, and the 0.5 terms in f_{blue} correspond to the probability that X falls within either Gaussian distribution, given that the color is blue.

Our objective is to compute the posterior probabilities of the label being red or blue, conditioned on observing the numerical value X. (Recall that X is derived from the two pdfs above once the class has been determined.)

Solution: We will solve this problem using a Bayesian approach,

$$\begin{aligned}
\Pr(C=\text{red}|X=x) &= \frac{f(C=\text{red},X=x)}{f(X=x)} && (1)\\
&= \frac{f(C=\text{red},X=x)}{f(C=\text{red},X=x)+f(C=\text{blue},X=x)} && (2)\\
&= \frac{\Pr(C=\text{red})f(X=x|C=\text{red})}{\Pr(C=\text{red})f(X=x|C=\text{red})+\Pr(C=\text{blue})f(X=x|C=\text{blue})} && (3)\\
&= \frac{0.5f(X=x|C=\text{red})}{0.5f(X=x|C=\text{red})+0.5f(X=x|C=\text{blue})} && (4)\\
&= \frac{f(X=x|C=\text{red})}{f(X=x|C=\text{red})+f(X=x|C=\text{blue})}, && (5)
\end{aligned}$$

where in (1) the class C is either red or blue, $f(\text{red},X)$ is the joint density of the label being red and the random variable X having the value x, and $f(X=x)$ is the marginal pdf for X having the value x;[1] the denominator of (2) partitions $f(X=x)$ into a sum of two conditional densities, $f(C=\text{red},X=x)$ and $f(C=\text{blue},X=x)$; (3) expresses joint

[1]A marginal pdf has the meaning that within all pairs $(C=c,X=c)$ we consider the density of $X=x$ irrespective of class value c.

densities as products of probabilities of the class and conditional densities of X given the class; (4) uses the Bayesian prior for the class probabiliies, $\Pr(\text{red}) = \Pr(\text{blue}) = 0.5$; and (5) is an algebraic simplification where the 0.5 terms all cancel out.

Next, recall that the two conditional densities were given to us,

$$f_{\text{blue}} = 0.5\mathcal{N}(-1,1) + 0.5\mathcal{N}(2,1) \quad \text{and} \quad f_{\text{red}} = \mathcal{N}(0,1).$$

These can be rewritten as follows,

$$f(X = x|C = \text{red}) = \frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(x-0)^2}{2\cdot 1}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \tag{6}$$

$$f(X = x|C = \text{blue}) = \frac{0.5}{\sqrt{2\pi \cdot 1}} e^{-\frac{(x-(-1))^2}{2\cdot 1}} + \frac{0.5}{\sqrt{2\pi \cdot 1}} e^{-\frac{(x-(2))^2}{2\cdot 1}} = \frac{0.5}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}} + \frac{0.5}{\sqrt{2\pi}} e^{-\frac{(x-2)^2}{2}}, \tag{7}$$

where we plugged the means μ and variances σ^2 into the standard pdf of a scalar Gaussian random variable, i.e., $f(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$.

We can now compute $\Pr(C = \text{red}|X = x)$ by plugging (6) and (7) into (5),

$$\begin{aligned} \Pr(C = \text{red}|X = x) &= \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + \frac{0.5}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}} + \frac{0.5}{\sqrt{2\pi}} e^{-\frac{(x-2)^2}{2}}} \\ &= \frac{e^{-\frac{x^2}{2}}}{e^{-\frac{x^2}{2}} + 0.5e^{-\frac{(x+1)^2}{2}} + 0.5e^{-\frac{(x-2)^2}{2}}}. \end{aligned} \tag{8}$$

We conclude the example by noting that the posterior probability for the class being blue satisfies

$$\Pr(C = \text{blue}|X = x) = 1 - \Pr(C = \text{red}|X = x),$$

implying that

$$\Pr(C = \text{blue}|X = x) = \frac{0.5e^{-\frac{(x+1)^2}{2}} + 0.5e^{-\frac{(x-2)^2}{2}}}{e^{-\frac{x^2}{2}} + 0.5e^{-\frac{(x+1)^2}{2}} + 0.5e^{-\frac{(x-2)^2}{2}}}.$$

This example can be modified quite easily. For example, each class could correspond to a conditional density for X having multiple Gaussian components; the variances of Gaussians could differ; and the prior probabilities for the two classes could also differ.