8 Optional

8.1 Throwing rocks: asymmetric loss

You just bought a new trebuchet and you are interested in making predictions about how far it can throw a rock. Your ballistics officer tells you that optimizing the squared error of your predictions is not appropriate for the problem. If your prediction is within some constant c of the true value then they can use your predicted value to aim the trebuchet such that it hits the castle, but if your prediction off by more than c, then using the prediction for aiming will cause the trebuchet to miss. However, the fact is, it's better for your prediction to be too short than too far.

So, we will let

$$L(a, g) = \begin{cases} 0 & \text{if } |a - g| < c \\ 1 & \text{if } a - g > c \\ 2 & \text{if } g - a > c \end{cases}$$

If you know that the range of the ball for these types of trebuchets is distributed as a Gaussian with mean μ and variance σ^2, what prediction minimizes loss L?

(This is a little bit tricky. It's fine to just write out an expression in terms of c, μ, and σ^2.)

Solution: The best prediction is the one that minimizes the expected loss over the randomness in a:

$$\begin{aligned} \hat{g} &= \arg\min_g \mathbb{E}_a[L(a, g)] \\ &= \arg\min_g \int_a L(a, g)p(a)da \\ &= \arg\min_g \int_{-\infty}^{\infty} L(a, g)\mathcal{N}(a|\mu, \sigma^2)da \\ &= \arg\min_g \int_{-\infty}^{g-c} L(a, g)\mathcal{N}(a|\mu, \sigma^2)da + \int_{g-c}^{g+c} L(a, g)\mathcal{N}(a|\mu, \sigma^2)da + \int_{g+c}^{\infty} L(a, g)\mathcal{N}(a|\mu, \sigma^2)da \end{aligned}$$

Note that we intentionally divided up the integral so that we can easily plug in values for $L(a, g)$. In the first integral, we have that the prediction is too far, and in the last integral, the prediction is too short:

$$= \arg\min_g \int_{-\infty}^{g-c} 2\mathcal{N}(a|\mu, \sigma^2)da + \int_{g+c}^{\infty} \mathcal{N}(a|\mu, \sigma^2)da. \quad (11)$$

We can now solve the optimization problem:

$$\begin{aligned} 0 &= \frac{1}{dg}\left[\int_{-\infty}^{g-c} 2\mathcal{N}(a|\mu, \sigma^2)da + \int_{g+c}^{\infty} \mathcal{N}(a|\mu, \sigma^2)da\right] \\ &= 2\mathcal{N}(g - c|\mu, \sigma^2) - \mathcal{N}(g + c|\mu, \sigma^2) \end{aligned}$$

which we see simplifies nicely by the fundamental theorem of calculus. We now solve for g:

$$2\mathcal{N}(g - c|\mu, \sigma^2) = \mathcal{N}(g + c|\mu, \sigma^2)$$

$$2\exp\left(-\frac{(g-c-\mu)^2}{2\sigma^2}\right) = \exp\left(-\frac{(g+c-\mu)^2}{2\sigma^2}\right)$$

We can take log of both sides and do some algebra to get the final answer:

$$g = \frac{-\sigma^2 \ln(2)}{2c} + \mu.$$

Intuitively, this makes sense. We slightly under-predict the mean because of our biased loss function.

8.2 Copy that: discrete Bayes update and decision theory

You have just bought a copy machine at a garage sale. You know it is one of two possible models, m_1 or m_2, but the tag has fallen off, so you're not sure which.

You do know that m_1 machines have a 0.1 "error" (bad copy) rate and m_2 machines have a 0.2 error rate.

(a) You use your machine to make 1000 copies, and 140 of them are bad. What is the maximum likelihood estimate of the machine's error rate? Explain why. (Remember that you're sure it's one of those two types of machines).

Solution: We first solve the MLE of the type of the machine, which we denote by $b \in \{1, 2\}$. Using a particular machine, the number of bad copies, denoted by k, is a random variable, as $k \sim \text{Binomial}(n, p_b)$. Thus,

$$\Pr(k \mid b) = \binom{n}{k} p_b^k (1 - p_b)^{n-k} \;\Rightarrow\; \log \Pr(k \mid b) = \log C + k \log p_b + (n-k)\log(1-p_b).$$

Here, C is the value of n *choose* k. With $n = 1000$, $k = 140$, $p_1 = 0.1$ and $p_2 = 0.2$, we have

$$\log \Pr(k \mid b = 1) = \log C + 140 \log(0.1) + 860 \log(0.9) = \log C - 412.97$$
$$\log \Pr(k \mid b = 2) = \log C + 140 \log(0.2) + 860 \log(0.8) = \log C - 417.22$$

We can see that $\log \Pr(k \mid b = 1) > \log \Pr(k \mid b = 2)$, which implies that the MLE of the type of the machine is $b_{\mathbf{ml}} = 1$. It follows that the machine's error rate is $p_{b_{\mathbf{ml}}} = 0.1$.

(b) Looking more closely, you can see part of the label, and so you think that, just based on the label it has a probability 0.2 of being an m_1 type machine and a probability 0.8 of being an m_2 type machine. If you take that to be your prior, and incorporate the data from part a, what is your posterior distribution on the type of the machine?