

Exercise: More reject options

In many applications, the classifier is allowed to “reject” a test example rather than classifying it into one of the classes. Consider, for example, a case in which the cost of a misclassification is \$10 but the cost of having a human manually make the decision is only \$3. We can formulate this as the following loss matrix:

Decision \hat{y}	true label y	
	0	1
predict 0	0	10
predict 1	10	0
reject	3	3

1. Suppose $P(y = 1|\mathbf{x})$ is predicted to be 0.2. Which decision minimizes the expected loss?
2. Now suppose $P(y = 1|\mathbf{x}) = 0.4$. Now which decision minimizes the expected loss?
3. Show that in general, for this loss matrix, but for any posterior distribution, there will be two thresholds θ_0 and θ_1 such that the optimal decision is to predict 0 if $p_1 < \theta_0$, reject if $\theta_0 \leq p_1 \leq \theta_1$, and predict 1 if $p_1 > \theta_1$ (where $p_1 = p(y = 1|\mathbf{x})$). What are these thresholds?