Machine Learning and Physics, Sheet 04

Jacob Waßelewsky, Johannes Schmidt, Dominik Geng

1 Bayes or Signal

With our parameters θ = gamma ray and $\bar{\theta}$ = background, we are given the priors $p(\theta) = 0.1$ and $p(\bar{\theta}) = 0.9$ as well as the likelihoods $p(x|\theta) = 0.95$ and $p(x|\bar{\theta}) = 0.1$. Therefore, we can calculate the posterior

$$\begin{aligned} p(\theta|x) &= \frac{p(x|\theta)p(\theta)}{p(x)} \\ &= \frac{p(x|\theta)p(\theta)}{p(x|\theta)p(\theta) + p(x|\bar{\theta})p(\bar{\theta})} = \frac{0.95 \cdot 0.1}{0.95 \cdot 0.1 + 0.1 \cdot 0.9} \approx 0.51. \end{aligned} \tag{1}$$

2 Bayes Classifiers

2.1 Binary Classification

Consider the Loss matrix The optimal Bayes classifier is given by minimizing the ex-

$L(y, \hat{y})$	$\hat{y} = 0$	$\hat{y} = 1$
$y = 0$	0	1
$y = 1$	10	0

Table 1: Asymmetric Loss matrix

pected loss

$$\bar{L} = \sum_{y^*} L(\hat{y}, y^*) p(y = y^*|x) \tag{2}$$

where we sum over all possible true labels y^*. This expands to

$$\bar{L} = (\hat{y}, 0)\, p(y = 0|x) + L(\hat{y}, 1)\, p(y = 1|x)$$

The individual loss matrix elements can be written as Kronecker deltas such that

$$\bar{L} = \delta_{\hat{y}1}\, p(y = 0|x) + 10\, \delta_{\hat{y}0}\, p(y = 1|x). \tag{3}$$

To minimize the loss, one chooses

$$\hat{y} = \arg\min_{\hat{y}} \bar{L} = \begin{cases} 0 & \text{if } \; 10\, p(y = 1|x) < p(y = 0|x) \\ 1 & \text{if } \; 10\, p(y = 1|x) > p(y = 0|x) \end{cases}, \tag{4}$$

which can be more conveniently written with the Heaviside step function

$$\hat{y} = \Theta\left(10\, p(y=1|x) - p(y=0|x)\right). \tag{5}$$

In words, we favor to label the event with $\hat{y} = 1$ and only assign class $\hat{y} = 0$ if the posterior for class 0 is more than 10 times higher than that of class 1. Such an asymmetric loss could be appropriate in the diagnostics of diseases. Here, it is more costly to assign an ill patient ($y = 1$) to the group of healthy ones ($y = 0$) (false negative) than the other way around (false positives). In this case, we want to be extra sure, that the patient is healthy before labeling him as such.

2.2 Reject option

Adding a reject class yields the Loss matrix We write the expected Loss as

$L(y,\hat{y})$	$\hat{y}=0$	$\hat{y}=1$	$\cdots$	$\hat{y}=k$	
$y=1$	α	0	1	$\cdots$	1
$y=2$	α	1	0	$\ddots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\ddots$	1
$y=k$	α	1	$\cdots$	1	0

Table 2: Reject option Loss matrix

$$\begin{aligned}
\bar{L} &= \sum_{y^*} L(\hat{y}, y^*) p(y = y^*|x) \\
&= \sum_{y^*} (\alpha\delta_{\hat{y}\,0} + 1 - \delta_{\hat{y}\,y^*}) p(y = y^*|x) \\
&= \alpha\delta_{\hat{y}\,0} + 1 - p(\hat{y}|x)
\end{aligned} \tag{6}$$

To minimize the loss, one chooses

$$\hat{y} = \arg\min_{\hat{y}} \bar{L} = \begin{cases} \arg\max_{y} p(y|x) & \text{if} \quad p(y|x) > 1 - \alpha \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$