DSE 220: Machine learning

Worksheet 6 — Solutions

1. *Inherent uncertainty.* This is somewhat subjective, but (b), (d) seem pretty clear-cut cases where perfect predictions are not possible.

2. *Logistic regression.* Since
$$\Pr(y = 1|x) = \frac{1}{1 + e^{-(w \cdot x + b)}},$$
we can rearrange terms to get
$$w \cdot x + b = \ln \frac{\Pr(y = 1|x)}{1 - \Pr(y = 1|x)}$$
   (a) $w \cdot x + b = \ln 1 = 0$
   (b) $w \cdot x + b = \ln 3$
   (c) $w \cdot x + b = -\ln 3$

3. *Form of the squashing function.*
$$\begin{aligned}
\Pr(y = 1|x) = \frac{\Pr(y = 1, x)}{\Pr(x)} &= \frac{\exp(-\|x - \mu_1\|^2/2\sigma^2)}{\exp(-\|x - \mu_1\|^2/2\sigma^2) + \exp(-\|x - \mu_2\|^2/2\sigma^2)} \\
&= \frac{1}{1 + \exp((\|x - \mu_1\|^2 - \|x - \mu_2\|^2)/2\sigma^2)} \\
&= \frac{1}{1 + \exp(2x \cdot (\mu_2 - \mu_1) + \|\mu_1\|^2 - \|\mu_2\|^2)}.
\end{aligned}$$
This is of the form $s(z)$ where $s(\cdot)$ is the squashing function and z is linear in x.

4. As the margin m increases, $f(m)$, the fraction of test points with margin $\geq m$, will decrease. We would expect $e(m)$, the error rate on points with margin $\geq m$, to decrease, but this might not happen.