Active Learning

Laura, Peter, Simon

We assume a prior $\pi(w)$ where $\mathbf{w}$ is a vector of parameters that describe our model. As the model we use a sigmoid function

$$\sigma\left(w_0, w_1, x\right) = \frac{1}{1 + \exp[-(w_0 - w_1 x)]} \tag{1}$$

NB: This might not be the best way to write the sigmoid. We collect data n data points by presenting a stimulus $x \in \mathbb{R}$ and observing a binary response $y \in \{0, 1\}$:

$$D = \{(x_1, y_1), \ldots, (x_n, y_n)\} \equiv (X, Y) \tag{2}$$

The likelihood of the parameters $\mathbf{w}$ given the data D is given by

$$\begin{aligned} P(Y \mid \mathbf{w}, X) &= \prod_{i=1}^{n} P(y_i \mid \mathbf{w}, x_i) && (3) \\ &= \prod_{i=1}^{n} \sigma\left(w_0 + w_1 x\right)^{y_i} \left(1 - \sigma\left(w_0 + w_1 x\right)\right)^{1-y_i} && (4) \end{aligned}$$

NB: The stimuli x are not considered part of the data, because we have control over it The posterior probability of the parameters $\mathbf{w}$ is

$$P(\mathbf{w} \mid X, Y) = \frac{P(Y \mid w, X)\pi(\mathbf{w})}{P(Y \mid X)} \tag{5}$$

The denominator, i.e. the marginal of the likelihood, is computed by taking the integral over all hypotheses:

$$\int P(Y \mid \mathbf{w}', X)\pi(\mathbf{w}')\,\mathrm{d}\mathbf{w}' = \iint P(Y \mid w_0, w_1, X)\pi(w_0, w_1)\,\mathrm{d}w_0'\,\mathrm{d}w_1' \tag{6}$$

The goal is to get a posterior $P(|X, Y)$ that is of low uncertainty. We use entropy as a measure of the current uncertainty of our estimation of $\mathbf{w}$. By *current* we mean that we use the data we have discovered in the n steps until now. To make this clear we write X_n, Y_n instead of X, Y:

$$H(\mathbf{w} \mid X_n Y_n) = -\int P(\mathbf{w}' \mid X_n, Y_n)\log[P(\mathbf{w}' \mid X_n, Y_n)]\,\mathrm{d}\mathbf{w}'\,. \tag{7}$$

In principle we would now like to choose our next stimulus x_{n+1} such that it minimizes the resulting entropy $H(\mathbf{w} \mid X_n, Y_n, x_{n+1}, y_{n+1})$, but we do not know what y_{n+1} is. So we want to find the x_{n+1} that minimizes the mean:

$$K(x_{n+1}) = H(\mathbf{w} \mid X_n, Y_n, x_{n+1}, y_{n+1} = 0)\ P(y_{n+1} = 0 \mid X_n, Y_n, x_{n+1}) \tag{8}$$

$$+\,H(\mathbf{w} \mid X_n, Y_n, x_{n+1}, y_{n+1} = 1)\ P(y_{n+1} = 1 \mid X_n, Y_n, x_{n+1})\,. \tag{9}$$

Here $P(y_{n+1} = 0/1 \mid X_n, Y_n, x_{n+1})$ is called the *predictive distribution*. Determining them again requires an integral over the hypotheses:

$$P(y_{n+1} = 0/1 \mid X_n, Y_n, x_{n+1}) = \int P(y_{n+1} = 0/1 \mid \mathbf{w}', x_{n+1})P(\mathbf{w}' \mid X_n, Y_n, x_{n+1}, y_{n+1} = 0/1)\,\mathrm{d}\mathbf{w}' \tag{10}$$