Two Pages on Logistic Regression

Mathias Winther Madsen

January 31, 2025

The inverse of the **logistic function**

$$\log \frac{p}{1-p} \qquad (0 < p < 1)$$

is the **sigmoid function**

$$\sigma(t) = \frac{1}{1+e^{-t}} = \frac{e^t}{e^t+1}$$
$$\sigma'(t) = \frac{e^{-t}}{(1+e^{-t})^2} = \sigma(t)\sigma(-t)$$

Its logarithm satisfies

$$\log \sigma(t) = \log\left(1+e^{-t}\right)$$
$$\log \sigma(t)' = \sigma(-t)$$

The sigmoid function is perfectly anti-symmetric around zero in the sense that

$$\underbrace{\sigma(t)}_{p} + \underbrace{\sigma(-t)}_{1-p} = 1.$$

It is therefore a good choice for relating a scalar evidence variable t to the probabilities of two exclusive possibilities, say, $y = +1$ or $y = -1$.

In logistic regression, we assume that the posterior log-odds in favor of some positive hypothesis is an affine function of some feature vector x:

$$u = \log \frac{p(+1 \,|\, x)}{p(-1 \,|\, x)} = a^T x + b.$$

This implies that

$$p(+1 \,|\, x) = \sigma(+u)$$
$$p(-1 \,|\, x) = \sigma(-u)$$

Hence if $y \in \{-1, +1\}$ is the true state of affairs, then $\sigma(yu)$ is the posterior probability assigned to the correct hypothesis. Under the logarithmic scoring rule for probability assignments, the model's loss at observing (x, y) is then

$$-\log \sigma(yu).$$

For notational convenience, we can define the vectors $w = (a\ b)$ and $z = y(x\ 1)$. We then get following formulas for the resulting loss function:

$$
\begin{aligned}
L(w) &= -\sum_n \log \sigma(w^T z_n) \\
\nabla_w L(w) &= -\sum_n \sigma(-w^T z_n) z_n \\
\nabla_w^2 L(w) &= +\sum_n \sigma(w^T z_n)\sigma(-w^T z_n) z_n z_n^T
\end{aligned}
$$

We can optimize this objective function using Newton's method, which here takes the form of the update rule

$$
w_{i+1} = w_i - \left(\nabla_w^2 L(w_i)\right)^{-1} \left(\nabla_w L(w_i)\right).
$$