Logistic Regression

- used for binary classification
- special case of [Generalized Linear Models]

fits the function

$$y = \frac{1}{1 + e^{-(\omega_0 x_0 + \omega_1 x_1 + \ldots)}} = \frac{1}{1 + e^{-\omega x}}$$

using the loss function:

$$L = \frac{-1}{N} \sum_{i=1}^{N} y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i))$$

Optimize weights w to minimize loss function using gradient descent.

Gradient derivation

Gradient of the sigmoid function

$$\begin{aligned}
\frac{dy}{d\omega} &= \frac{(1 + e^{-\omega x})(0) - (1)(-xe^{-\omega x})}{(1 + e^{-\omega x})^2} \\
&= \frac{xe^{-\omega x}}{(1 + e^{-\omega x})^2} \\
&= x \frac{1 - 1 + e^{-\omega x}}{(1 + e^{-\omega x})^2} \\
&= x \left(\frac{1 + e^{-\omega x}}{(1 + e^{-\omega x})^2} - \frac{1}{(1 + e^{-\omega x})^2} \right) \\
&= x \left(\frac{1}{1 + e^{-\omega x}} - \frac{1}{(1 + e^{-\omega x})^2} \right) \\
&= x \left(\frac{1}{1 + e^{-\omega x}} \left(1 - \frac{1}{1 + e^{-\omega x}} \right) \right) \\
&= xy(x)[1 - y(x)]
\end{aligned}$$

gradient of the loss function

For conciseness, I'll refer to the ground truth as y_t and the predicted value y_p

$$
\begin{aligned}
\frac{\partial L}{\partial \omega} &= \frac{\partial}{\partial \omega}(-y_t \ln(y_p) - (1-y_t)\ln(1-y_p)) \\
&= \frac{-y_t}{y_p}\frac{\partial y_p}{\partial \omega} - \frac{1-y_t}{1-y_p}\frac{\partial (1-y_p)}{\partial \omega} \\
&= \frac{-y_t}{y_p}\frac{\partial y_p}{\partial \omega} + \frac{1-y_t}{1-y_p}\frac{\partial y_p}{\partial \omega} \\
&= \frac{\partial y_p}{\partial \omega}\left(\frac{-y_t}{y_p} + \frac{1-y_t}{1-y_p}\right) \\
&= \frac{\partial y_p}{\partial \omega}\left(\frac{-y_t(1-y_p) + (1-y_t)(y_p)}{y_p(1-y_p)}\right) \\
&= xy_p(1-y_p)\left(\frac{-y_t + y_t y_p + y_p - y_t y_p}{y_p(1-y_p)}\right) \\
&= x(y_p - y_t)
\end{aligned}
$$

References

- https://medium.com/analytics-vidhya/derivative-of-log-loss-function-for-logistic-regression-9b832f025c2d