1 Logistic Regression

Logistic regression makes use of the logistic, or sigmoid, function,

\[ \sigma(x) = \frac{1}{1 + e^{-x}} \]

This function helps us model the probability of Bernoulli random variables, i.e. binary categorical data, as it maps from \( \mathbb{R} \to [0, 1] \). With the category labels \( y \in \{-1, 1\} \), we define the data likelihood for \( y = 1 \) as

\[ p(y = 1 | x, \beta) = \sigma(\beta^T x) = \frac{1}{1 + e^{-\beta^T x}} \]

And more generally,

\[ p(y | x, \beta) = \frac{1}{1 + e^{-y \beta^T x}} \]

1. Show that \( \beta^T x \) is interpreted as the log odds for \( y = 1 \), \( \ln \frac{p(y = 1 | x, \beta)}{p(y = -1 | x, \beta)} \)
2. Write down the likelihood for data \( Y = \{ y \}_i^{n}, X = \{ x \}_i^{n} \)

3. Suppose we wanted to find the \( \beta \) that maximized the data likelihood. Which of the following is this equivalent to? Identify all that apply.

   a) The \( \beta \) minimizing the log likelihood
   b) The \( \beta \) maximizing the negative likelihood
   c) The \( \beta \) minimizing the negative log likelihood
   d) The \( \beta \) maximizing the negative log likelihood

4. Write down the expression for the choice above.

5. The quantity \( \ln(1 + e^{-y\beta^T x}) \) is also known as the logistic loss. For \( y = 1 \) and \( y = -1 \), sketch this loss as a function of \( \beta^T x \). (Try sketching \( 1 + e^{-y\beta^T x} \) first.)

6. How would you alter the logistic loss to include an \( l_2 \) regularization term?
2 Computational Example

Follow along with the notebook available at: https://nbviewer.jupyter.org/github/justmarkham/DAT8/blob/master/notebooks/12_logistic_regression.ipynb

Description of the dataset: https://archive.ics.uci.edu/ml/datasets/Glass+Identification

Material from Kevin Markham.

For TAs:

1. Sections 2,3, and 6 are most relevant. Cover the others as you want to.

2. Note that for logistic regression scikit-learn uses $l_2$ regularization by default with $c = 1$!