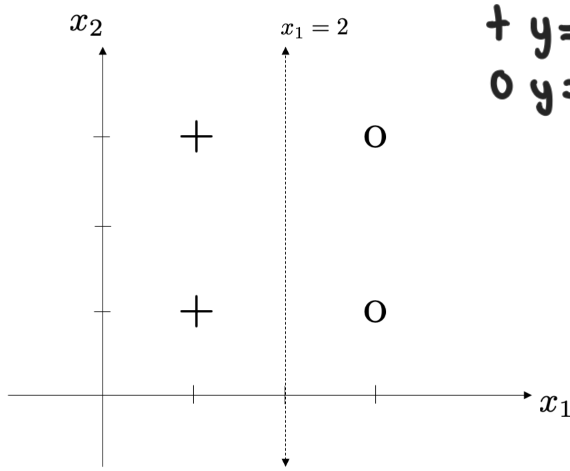
Logistic Regression

Submit a PDF of your answers to Canvas

1. Consider the four data points shown in the below. The data points at $(1,1)$ and $(1,3)$ belong to the class labeled $y = 1$, while the data points at $(3,1)$ and $(3,3)$ belong to the class labeled $y = -1$.



$+$ $y = 1$
0 $y = -1$

a) A decision boundary at $x_1 = 2$ can be expressed as the set of points that satisfy $\boldsymbol{x}^T \boldsymbol{w} = 0$, where

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} W_0 \\ -W_0/2 \\ 0 \end{bmatrix}$$

Find w_0, w_1, w_2 so that the set of points that satisfy $\boldsymbol{x}^T \boldsymbol{w} = 0$ correspond to the vertical line at $x_1 = 2$. Note that your answer will not be unique. Express your answer for w_1 in terms of w_0.

b) The starting point for motivating logistic regression is to assume that $y \in \{-1, 1\}$ is a Bernoulli random variable with bias that depends on $\boldsymbol{x}$:

$$p(y = 1|\boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{x}^T \boldsymbol{w}}}$$

$$X^T W = W_0 \left(1 - \frac{X}{2}\right)$$

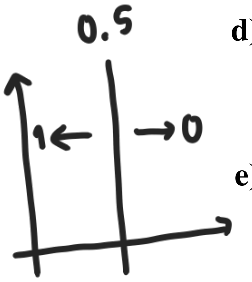
and

$$p(y = -1|\boldsymbol{x}) = \frac{1}{1 + e^{\boldsymbol{x}^T \boldsymbol{w}}}$$

Find an expression for the likelihood $L(\boldsymbol{w}) = p(y_1, y_2, y_3, y_4)$ given the data in the figure, and assuming (i) independence and (ii) that $\boldsymbol{w}$ is restricted to correspond to the result from (a).

$$L(W) = \left(\frac{1}{1 + e^{-W_0/2}} \right)^4$$

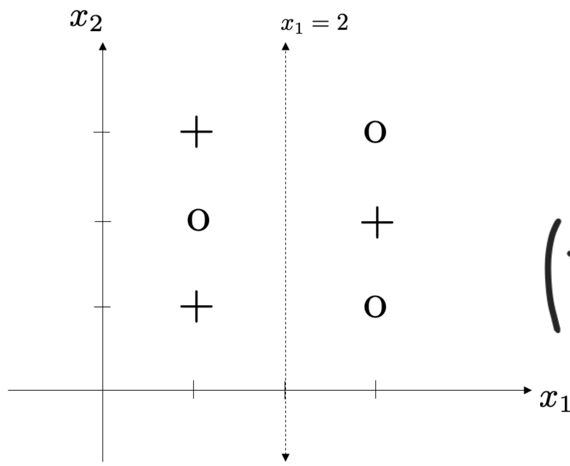
c) Find an expression for the negative log-likelihood. $4\log(1+e^{-W_0/2})$



d) Recall that we are usually interested in maximizing the likelihood $L(\boldsymbol{w})$, or equivalently minimizing the negative log-likelihood. Does the likelihood have a maximum value in this case? No (larger w_0 → smaller denominator → larger $L(w)$)

e) Sketch or describe the estimate of $p(y=1|\boldsymbol{x})$ for a large w_0 over the x_1, x_2 plane.

2. You collect two more labeled data points, which are show on the plot below. The points are at $(1,2)$ and $(3,2)$.



$$\left(\frac{1}{1+e^{-W_0/2}}\right)^4 \left(\frac{1}{1+e^{W_0/2}}\right)^2$$

a) As before, find an expression for the likelihood $L(\boldsymbol{w})$, under the restriction that the decision boundary is a vertical line at $x_1 = 2$.

b) Maximize the likelihood function to specify $\boldsymbol{w}$ (note: you may need to use a computer to find an explicit answer, alternatively, you may leave your answer in a simple, but not explicit, analytic form).

c) Sketch or describe the estimate of $p(y=1|\boldsymbol{x})$ over the x_1, x_2 plane, and describe the effect of changing w_0 on the surface. How does this compare to your result from problem 1? There's a fixed value of W_0 that minimizes $L(w$

$$\frac{d}{dx}L(x) = \frac{-(e^{x/2}-2)e^{2x}}{(e^{x/2}+1)e^{7}} \rightarrow \text{root at } x = 2\ln(2)$$

$\therefore W_0 = 1/2$ maximizes likelihood