1. （2pt）Consider applying logistic regression to the following dataset:

x_1	x_2	y
0	0	1
0	1	0
1	0	0
1	1	1

a) (1pt) If we use raw feature x_1 and x_2, the model is

$$p(y=1|\boldsymbol{x},\boldsymbol{w}) = \text{sigmoid}\,(w_0+w_1x_1+w_2x_2)$$

What is the minimum achievable training error (= $\frac{\#\,incorrect}{\#\,total}$) in this case? Give weights that achieve the minimum error.

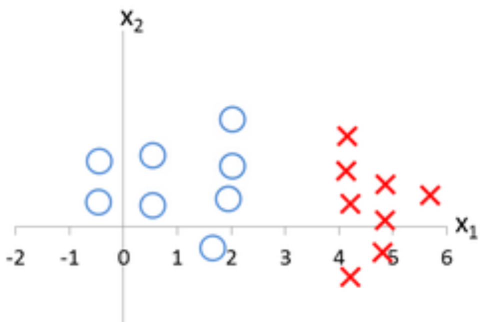
b) (1pt) Next consider using an additional feature x_1x_2 in addition to the raw feature x_1 and x_2. The model now is

$$p(y=1|\boldsymbol{x},\boldsymbol{w}) = \text{sigmoid}\,(w_0+w_1x_1+w_2x_2+w_3x_1x_2)$$

What is the minimum achievable training error in this case? Give weights that achieve the minimum error.

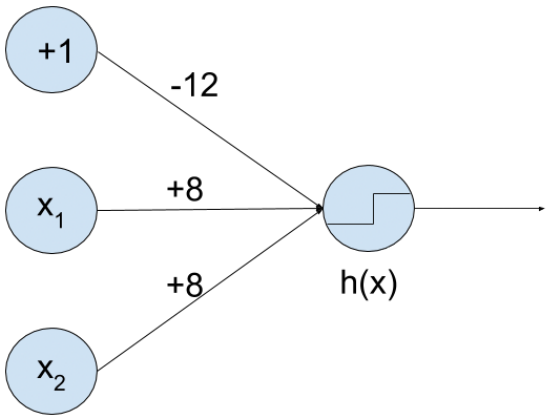
2. (2pt) Consider the binary classification problem, in which each observation $x^{(l)}$ is known to belong to one of two classes, corresponding to $r = 0$ and $r = 1$. Suppose that the procedure for collecting training data is imperfect, so that training points are sometimes mislabelled. For every data point $x^{(l)}$, instead of having a value $r^{(l)} \in \{0,1\}$ for the class label, we have instead a value $\pi^{(l)}$ representing the probability that $r^{(l)} = 1$. Given a logistic regression model y=p(r = 1|$\boldsymbol{x}$, $\boldsymbol{w}$), write down the loss function appropriate to such a data set.

3. (1pt) Consider the training set below, where “x” denotes positive examples (y=1) and “o” denotes negative examples (y=0). Suppose you train an SVM (which will predict 1 when $w_0+w_1x_1+w_2x_2 \geqslant 0$). What values might the SVM give for w_0, w_1, and w_2?



(a) $w_0=3, w_1=1, w_2=0$
(b) $w_0= -3, w_1=1, w_2=0$
(c) $w_0=3, w_1=0, w_2=1$
(d) $w_0= -3, w_1=0, w_2=1$

4. (1pt) You are given the following neural networks which take two binary valued inputs x_1 , $x_2 \in \{0, 1\}$ and the activation function is the threshold function($h(x) = 1$ if $x > 0$; 0 otherwise). Which of the following logical functions does it compute?



(a) OR (b) AND (c) NAND (d) None of the above.

5. (4pt) We have a function which takes a two-dimensional input x = (x_1 , x_2) and has two parameters w = (w_1,w_2) given by $f(x,w) = \sigma(\sigma(x_1w_1)w_2 + x_2)$ where $\sigma(x) = 1/(1+e^{-x})$. We use backpropagation to estimate the right parameter values. We start by setting both the parameters to 0. Assume that we are given a training point $x_1 = 1$, $x_2 = 0$, $y = 5$ and minimize the MSE loss.

   1)(2pt) What is the value of $\partial f/\partial w_2$? Show the computation process.

   2) (2pt) If the learning rate is 0.5, what will be the value of w_2 after one update using backpropagation algorithm? Show the computation process