Logistic Regression Quiz

1. Logistic regression is a method for classification.

2. How many classes can a logistic regression model choose between? Two.

3. In a logistic regression model, we learn a model of
   - The probability that an instance is in the positive class
   - A straight line decision boundary

4. In the model formulation,

$$p(y=1 \mid \underline{x}) = \sigma\left(\sum_{j=0}^{D} w_j x_j\right) = \sigma(\underline{w}^T \underline{x})$$

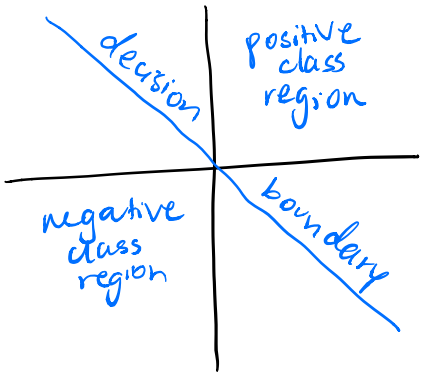
   w and x are:

$$\underline{w} = (w_0, w_1, \ldots, w_p)^T$$
$$\underline{x} = (1, x_1, x_2, \ldots, x_D)^T$$

5. The purpose of the sigmoid function $\sigma(w^T x)$ in the logistic regression model is:
   - to produce values that sum up to 1
   - to produce values that can be used as probabilities

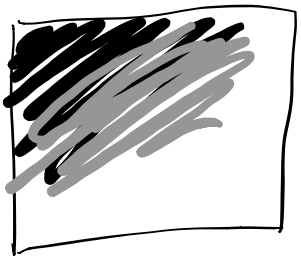
6. We have two two-dimensional datasets D1 and D2. We use logistic to learn decision boundaries, and find that they are W1 = [0 1 1] and W2 = [0 2 2]. For any new datapoint, the two models will classify the datapoint into the same class. (W1 and W2 are scalar equivalents)

7. For dataset D2 in the previous question, which of the following diagrams indicate the orientation of the decision boundary and how it distinguishes between positive and negative instances?



8. Assume the same problem setting as the previous two questions and label the model D1 with M1 and D2 with M2. In the following diagrams set in X1,X2 space with the horizontal axis being X1, white represents certainty of one class, black represents certainty of the other class, and greys represent degree of uncertainty. You can see there is a smooth transition from certainty of one class to the other. Which pair of diagrams represent better the change in probability from one class to the other as the new datapoint moves away from and towards the decision boundary?

Note: consider what happens to $p(y=1|w,x) = \sigma(w^T x)$ when there is a small change in $\mathbf{x}$. What is the effect of the parameter values? For the same small change in $\mathbf{x}$, does $p(y=1|w,x)$ change for more or less when the parameter values are larger?



9. We learn the logistic regression model using a numerical optimisation method -- solving an equation that sets the derivative of the log likelihood to zero.

10. We can use feature transforms in Logistic Regression (such as using RBF on Linear Regression).

11. Facts:
    - Logistic Regression uses a discriminative approach
    - Naive Bayes uses a generative approach
    - Logistic regression models p(y|x), where y is the output and x is the input

12. If we are building a classifier for K > 2 classes, which of the following are true of the approach we use?
    - ○ We build K logistic regression models
    - ○ We use the softmax function to combine the parameters produced by the K models
    - ○ The softmax function combines the output of the exponential function

- Each model Mi (i = 1,...,K) learns a set of parameters which enable it to decide, for a new instance, if should be classified into class Ci or into one of the other classes