1. Note that the answer to this question is not at all obvious. You’ll need to struggle with it a bit and do some reading. However, the insight you’ll gain will be extremely useful as we move on to other algorithms.

Consider the following learning algorithm known as the dual perceptron algorithm:

\begin{align*}
\text{Let } & \alpha_i = 0 \text{ for } i = 0, \ldots, N \\
\text{Repeat} & \text{ forever} \\
\text{Accept} & \text{ training example } (x_i, y_i) \\
& \text{if } \sum_l \alpha_l x_l y_l y_i \leq 0 \\
& \quad \alpha_i = \alpha_i + 1
\end{align*}

Note that \( \alpha_i \) is a counter of the number of times training example \( x_i \) has been misclassified. The outer loop that repeats forever cycles through the \( N \) training instances.

Explain intuitively what this algorithm does and why it is sensible. You can use the web to help you answer this question. If you do, cite the sources you used. Note that this algorithm is equivalent to the online perceptron algorithm with learning rate 1 and weight vector \( w = \sum_l \alpha_l x_l y_i \).

2. Show that for a linearly separable dataset, the maximum likelihood solution for the logisitic regression model is obtained by finding a weight vector \( w \) whose decision boundary \( w \cdot x = 0 \) separates the classes and then taking the magnitude of \( w \) to infinity.

3. Get the Iris dataset from the UCI Machine Learning Repository:


At the top of that page there are links for the data and a description of the data. Implement Linear Discriminant Analysis just for this dataset. That is, your implementation does not need to work for any dataset with any dimensions, it just needs to work for this one dataset.

Turn in your code and, for each instance, the value of \( \log_2(p(x, y)) \) for each of the 3 classes in the Iris dataset. Note the use of the log function is meant to make the output more readable by the TA. Be sure to use the base 2 log or your numbers will all look wrong.