Q3.

(a) we want to minimize

$$L(\vec{w}) = -\sum_{i=1}^{N} y_i \log P(C_1|x_i) + (1-y_i) \log(1-P(C_1|x_i))$$

$$P(C_1|x_i) = \frac{1}{1+e^{-w^T \vec{x}_i}}$$

let C_1 is the class of orange, C_2 not.

and

$\vec{w} = [w_1, w_2, b]$

$\vec{x}_1 = [4, 4, 1]$ $y_1 = 1$

$\vec{x}_2 = [6, 4, 1]$ $y_2 = 1$

$\vec{x}_3 = [6, 5, 1]$ and $y_3 = 1$

$\vec{x}_4 = [6, 8, 1]$ $y_4 = 0$

$\vec{x}_5 = [6, 10, 1]$ $y_5 = 0$

$\vec{x}_6 = [8, 8, 1]$ $y_6 = 1$

$\vec{x}_7 = [8, 10, 1]$ $y_7 = 0$

(b)

$$\frac{\partial L(\vec{w})}{\partial w} = -\sum_{i=1}^{7} (y_i - P_i) \cdot \vec{x}_i$$

$$P_i = \frac{1}{1-e^{-\vec{w}^T \vec{x}_i}}$$

$$W_{i+1} = W_i + \lambda \frac{\partial L(\vec{w})}{\partial w}$$

Initially: $\vec{w}_0 = [0.3, -0.2, 0.7]$

accuracy $= \frac{4}{7}$

Iteration 1: $\vec{w}_1 = [0.20477, -0.35426, 0.68685]$

accuracy $= \frac{6}{7}$

Iteration 2: $\vec{w}_2 = [0.27913, -0.30894, 0.69952]$

accuracy $= \frac{7}{7}$

Iteration 3: $\vec{w}_3 = [0.27208, -0.35574, 0.69979]$

accuracy $= \frac{7}{7}$

(c)

(3,3) is orange

(4,10) is not orange

(9,8,1) is orange

(9,10,1) is Orange

(d)

Advantage:

Logistic regression models are highly interpretable. The coeeficientes of the model give a direct indication of the importance and direction of the feature on the probability being modeled.

Disadvantage:

It only works on a linear boundry.