Section 7

1. According to $\hat{Y} = \begin{bmatrix} 0.6 \\ 0.7 \\ 0.4 \end{bmatrix}$, $\because 0.6 > 0.5$, then $\hat{y}_1 = 1 = y_1$

$0.7 > 0.5$, then $\hat{y}_2 = 1 = y_2$

$0.4 < 0.5$, then $\hat{y}_3 = -1 = y_3$

$\therefore$ then accuracy is 100%.

2. No. Because $\hat{Y}_3 = 0.4$, which is not very smaller than 0.5, i.e. the confidence to classify x_3 as a negative sample is still small. We expect $\hat{Y}_3 \to 0$.

3. The output of Logistic Regression is $[0,1]$, which can be used as probability to identify different sample's confidence / score.

4. 1) For classes i and j. their decision boundary $y_i(x) = y_j(x)$, i.e.

$$(\theta_i - \theta_j)^T x + (\theta_{i0} - \theta_{j0}) = 0$$

If $\theta_i \approx \theta_j$, $\theta_{i0} \approx \theta_{j0}$, then any $\forall x \in \mathbb{R}^n$, $(\theta_i - \theta_j)^T x_n + (\theta_{i0} - \theta_{j0})$ will close to 0.
i.e. lies on the boundary

2) $L = L_0 \cdot \frac{1}{\|\theta_i - \theta_j\|}$

5. 1) If $x_i \in$ class k, then $y_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$, where $y_{ik} = 1$, otherwise 0.

2) Because: ① the scales of features might be different,
② The ground truth label vector has length 1, we need to scale our prediction labels for better comparing

6. No. If $\theta^T x > 0$, $y = 0$, $\theta^T x < 0$, $y = 1$, the prediction is not continuous.