<u>Homework 2:</u>

A.

$$P(y|x;W) = \frac{e^{w_y^T * x}}{\sum_{c=1}^{C} e^{w_y^T * x}}$$

Take the negative log of a single instance and simplify

$$-\ln(P(y|x;W)) = -\ln\left(\frac{e^{w_y^T * x}}{\sum_{c=1}^{C} e^{w_y^T * x}}\right)$$
$$= \ln\left(\sum_{c=1}^{C} e^{w_y^T * x}\right) - \ln\left(e^{w_y^T * x}\right)$$
$$= \ln\left(\sum_{c=1}^{C} e^{w_y^T * x}\right) - w_y^T * x$$

Take the summation for n data points and simplify

$$L(D) = \sum_{i=1}^{n} \left(\ln\left(\sum_{c=1}^{C} e^{w_{y_i}^T * x_i}\right) - w_{y_i}^T * x_i\right)$$
$$L(D) = \sum_{i=1}^{n} \ln\left(\sum_{c=1}^{C} e^{w_{y_i}^T * x_i}\right) - \sum_{i=1}^{n} w_{y_i}^T * x_i$$

B.

Derive the gradient of the logistic regression likelihood with respect to any one of the W_c vectors

$$\Delta w_c \;\; L = \frac{d}{dwc} L(D))$$
$$= \frac{d}{dw_c}(\sum_{i=1}^{n} \quad (ln(\sum_{c\prime} \quad (e^{w_{c\prime}^T x_i})) - \sum_{i=1}^{n} \quad (w_{y_i}^T x_i))$$
$$= \frac{d}{dwc}(\sum_{i=1}^{n} \quad (ln(\sum_{c\prime} \quad (e^{w_{c\prime}^T x_i}) - w_{y_i}^T x_i)))$$
$$= \sum_{i=1}^{n} \quad (\frac{d}{dw_c}(ln(\sum_{c\prime} \quad (e^{w_{c\prime}^T x_i})) - \frac{d}{dw_c}(w_{y_i}^T x_i))$$
$$= \sum_{i=1}^{n} x_i(\frac{\sum_{c\prime} \quad (e^{w_{c\prime}^T x_i} * I(c\prime = c) x_i)}{\sum_{c\prime}^{C} \quad (e^{w_{c\prime}^T x_i})} - I(y_i = c))$$

$= \sum_{i=1}^{n} \quad x_i(\frac{\sum_{c\prime} \quad (e^{w_{c\prime}^T x_i} * I(c\prime = c))}{\sum_{c\prime}^{C} \quad (e^{w_{c\prime}^T x_i})} - I(y_i = c))$

$= \sum_{i=1}^{n} \quad x_i(\frac{e^{w_{c\prime}^T x_i}}{\sum_{c\prime} \quad (e^{w_{c\prime}^T x_i})} - I(y_i = c))$

C.

$$P(y|x;W) = \frac{e^{w_y^T * x}}{\sum_{c=1}^{C} e^{w_y^T * x}}$$

$= \frac{exp(0)}{\sum_{c=1}^{C} \quad (exp(0))}$

$= \frac{1}{C}$

Explanation of why this probability is reasonable as the maximally regularized solution:

This probability make sense because it eliminate all biases in the data. All vector start with the same weight give all the classes the same chance of being the solution.