Maximum Negative Liklihood, J_i, for Logistic Regression

$$J_j(\theta_i) = -y^{(j)} \log \sigma(x^{(j)} \theta_i^T) - (1-y^{(j)}) \log(1-\sigma(x^{(j)}\theta_i^T))$$

J = cost (gradient) function θ = weights (parameters)

σ = sigmoid function - threshold function which linear response passed to

j = current observation, i = current parameter, θ, x = feature data, y = label data

① Gradient, $\frac{\partial J}{\partial \theta}$, of Maximum Negative Liklihood

① Given there are 3 parameters: θ_0, θ_1, and θ_2, the gradient vector would be

$$\nabla J_j = \left\langle \frac{\partial J_j}{\partial \theta_0}, \frac{\partial J_j}{\partial \theta_1}, \frac{\partial J_j}{\partial \theta_2} \right\rangle$$

Solving for the i^{th} component, $\frac{\partial J_j}{\partial \theta_i}$, of ∇J of the j^{th} data point $(x = x^{(j)})$

$$J(\theta_i) = -y \log \sigma(x \cdot \theta_i^T) - (1-y) \log(1-\sigma(x \cdot \theta_i^T))$$

$$\partial J(\theta_i)_j = -\frac{\partial}{\partial \theta_i} y \log \sigma(x \cdot \theta_i^T) + \frac{\partial}{\partial \theta_i} (1-y) \log(1-\sigma(x \cdot \theta_i^T))$$

$$= \left[-\frac{y}{\sigma(x \cdot \theta_i^T)} + \frac{(1-y)}{1-\sigma(x \cdot \theta_i^T)} \right] \frac{\partial}{\partial \theta_i} \sigma(x \cdot \theta^T)$$

$$= \left[\frac{-y}{\sigma(x \cdot \theta_i^T)} + \frac{(1-y)}{1-\sigma(x \cdot \theta_i^T)} \right] \sigma(x \cdot \theta_i^T) \left[1-\sigma(x \cdot \theta_i^T)\right] x$$

$$= \left[\frac{-y + \sigma(x \cdot \theta_i^T)}{\sigma(x \cdot \theta_i^T)(1-\sigma(x \cdot \theta_i^T))} \right] \sigma(x \cdot \theta_i^T) \left[1-\sigma(x \cdot \theta_i^T)\right] x$$

$$= \left[\sigma(x \cdot \theta_i^T) - y\right] x^{(j)} \Rightarrow \boxed{\nabla_\theta J = (\hat{y} - y) x}$$