HW3 - DVA

mmendiola3

1. Theory

a. Write down the formula for computing the gradient of the loss function used in Logistic Regression. Specify what each variable represents in the equation.

Cost function:

$$\theta_{MLE} = argmin_\theta \sum_{i=1}^{n} log(1 + \exp(y^i < \theta, x^i >)$$

Update function:

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} \sum_{i=1}^{n} log(1 + \exp(y^i < \theta, x^i >))$$

$$= \theta_j - \alpha \sum_{i=1}^{n} \frac{1}{1 + \exp(y^i < \theta, x^i >)} \cdot \frac{\partial}{\partial \theta_j} \exp(y^i < \theta, x^i >)$$

$$= \theta_j - \alpha \sum_{i=1}^{n} \frac{\exp(y^i < \theta, x^i >) \cdot \frac{\partial}{\partial \theta_j} y^i < \theta, x^i >}{1 + \exp(y^i < \theta, x^i >)}$$

$$= \theta_j - \alpha \sum_{i=1}^{n} \frac{\exp(y^i < \theta, x^i >) \cdot y^i x_j^i}{1 + \exp(y^i < \theta, x^i >)}$$

$$= \theta_j - \alpha \sum_{i=1}^{n} \frac{y^i x^i}{1 + \exp(-y^i < \theta, x^i >)}$$

note: $< \theta, x^i >$ is constant with the exception of $\theta_j \cdot x_j^i$

Terms:

- θ_j : The value of the current parameter vector at feature index j
- α : The learning rate, which decreases over each training iteration
- n : The number of training samples
- y^i :