Q3 → Extra Credit

In imbalanced classes, weighting the classes implies multiplying the occurences of the minority class by the r, where $r = \frac{\# \text{ Majority class}}{\# \text{ Minority Class}}$

For logistic regression the likelihood function is

$$L(\theta) = \prod_{i=1}^{m} P(y^i \mid x^i ; \theta) = \prod_{i=1}^{m} h_\theta(x^i)^{y^i} \cdot (1 - h_\theta(x^i))^{1-y^i}$$

When introducing the weighting it transforms into

$$\Rightarrow \prod_{i=1}^{m} h_\theta(x^i)^{y^i} \cdot \left((1 - h_\theta(x^i))^{1-y^i}\right)^{\underline{\underline{r}}}$$

Note the introduction of an additional exponent for the negative class. The negative class is assumed to be in the minority

Calculating Log likelihood

$$\text{Log}(L(\theta)) = \sum_{i=1}^{m} y^i \cdot \log h_\theta(x^i) + \underline{\underline{r}} \cdot (1-y^i) \cdot \log(1 - h_\theta(x^i))$$

Taking the gradient

$$\nabla_\theta \text{Log}(L(\theta)) = \sum_{i=1}^{m} \frac{y^i}{h_\theta(x^i)} \cdot x^i \cdot h_\theta(x^i) \cdot 1 - h_\theta(x^i))$$

$$+ r \cdot \frac{(y^i - 1)}{1 - h_\theta(x^i)} \cdot x \cdot h_\theta(x^i) \cdot (1 - h_\theta(x))$$

$$\Rightarrow \sum_{i=1}^{m} y \cdot x(1 - h_\theta(x)) + r \cdot (y-1) \cdot x \cdot h_\theta(x)$$

$$\sum_{i=1}^{m} x\left[y - y \cdot h_\theta(x) + r y \cdot h_\theta(x) - r\, h_\theta(x)\right]$$

$$\sum_{i=1}^{m} x\left[y - y \cdot h_\theta(x)(1 - r) - r\, h_\theta(x)\right]$$

$$= \sum_{i}^{m} x\left[y\left(1-h_\phi(x)(1-r)\right) - r\, h_\phi(x)\right]$$

Therefore derivative wrt w or ϕ

$$\frac{\partial}{\partial \phi_j} \log(L(\phi)) = \sum x\left[y(1-h_\phi(x)(1-r)) - r\, h_\phi(x)\right]$$

Derivative wrt b or ϕ_0

$$= \frac{\partial}{\partial \phi_0} \log(L(\phi)) = \sum_{i=1}^{m} \left[y(1-h_\phi(x)(1-r)) - r\, h_\phi(x)\right]$$

3b) Adding L2 Norm compoment to the equation

$$-\log(L(\phi)) + \frac{\alpha}{2}\|\phi\|^2 \qquad \alpha = \text{regularization weight}$$

The Equation ~~max or less~~ remains constant for the Constant "b" as "b" is not regularized

For the weights

$$\frac{\partial}{\partial \phi_j} \log(L(\phi)) = \sum_{j=1}^{m} x\left[y(1-h_\phi(x)(1-r)) - r\, h_\phi(x)\right] + \cancel{2}\cdot\frac{\alpha}{\cancel{2}}\cdot\phi_j$$

$$= \sum_{j=1}^{m} x\left[y(1-h_\phi(x)(1-r)) - r\, h_\phi(x)\right] + \phi_j$$