→) Logistic regression

Dataset:

.) m = no of training samples
.) n = no of features

→) Equations:

.) i refers to one training sample meaning 1 row

.) $x_i = [x_1, x_2 \cdots x_n] \rightarrow (1 \times n)$

.) $W = [w_1, w_2 \cdots w_n] \rightarrow (1 \times n)$

.) $B = [B] \rightarrow (1 \times 1)$

$\rightarrow (1 \times 1)$

.) $Z = W @ x_i.T + B$

.) $\hat{y} = \frac{1}{1+e^{-z}}$

.) $L+= -\frac{1}{m}[y \log(\hat{y}) + (1-y)\log(1-\hat{y})]$

.) $\frac{\partial L+}{\partial w} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z} \times \frac{\partial z}{\partial w}$

$= -\frac{1}{m}\left[\frac{y}{\hat{y}} - \frac{1-y}{(1-\hat{y})}\right] \times \frac{e^{-z}}{(1+e^{-z})^2} \times x_i$

$$= -\frac{1}{m}\left[\frac{y - y\vec{y} - \vec{y} + y\vec{y}}{\vec{y}(1-\vec{y})}\right] \times \left[\frac{1}{1+e^{-z}} - \frac{1}{(1+e^{-z})^2}\right] \times x_i$$

$$= -\frac{1}{m}\left[\frac{y - \vec{y}}{\vec{y}(1-\vec{y})}\right] \times \left[\vec{y} - \vec{y}^2\right] \times x_i$$

$$= -\frac{1}{m}(y - \vec{y})\, x_i$$

$$\text{+= } \underset{(1\times 1)}{\frac{1}{m}(\vec{y} - y)} \; @ \; \underset{\rightarrow (1\times n)}{x_i}$$

.) $\frac{\partial L}{\partial B} \text{+=} \frac{\partial L}{\partial \vec{y}} \times \frac{\partial \vec{y}}{\partial z} \times \frac{\partial z}{\partial B}$

$$\text{+= } \frac{1}{m}(\vec{y} - y)$$

.) $W = W - \alpha \frac{\partial L}{\partial W}$

.) $B = B - \alpha \frac{\partial L}{\partial B}$

Performance metrics:

.) TP (True Positive): $\vec{y} = 1$, $y = 1$
.) FP (False Positive): $\vec{y} = 1$, $y = 0$
.) TN (True Negative): $\vec{y} = 0$, $y = 0$
.) FN (False Negative): $\vec{y} = 0$, $y = 1$

.) Classification accuracy = $\frac{TP+TN}{TP+FP+FN+TN}$

→) Total correct predictions / total predictions

.) Misclassification rate = $\frac{FP+FN}{TP+FP+FN+TN}$

→) Total wrong predictions / total predictions

→) 1 − classification accuracy

.) Precision = $\frac{TP}{TP+FP}$

→) out of all true predictions how many were actually right

.) Recall = $\frac{TP}{TP+FN}$

→) out of all actual true values how many were predicted true

.) F1-Score = $\frac{2 * recall * precision}{recall + precision}$

Findings:

1.) Better to initialize weights with some values

2.) Better to normalize integer values >100 or causes overflow

3.) epochs = 100 accuracy = 57%
epochs = 200 accuracy = 58.7%
epoch = 300 accuracy = 59.7%
epoch = 400 accuracy = 60.7%
epoch = 500 accuracy = 61.8%

Seems directly ∝, ended at 800

4.) accuracy = 62.8%
precision = 37.8%
Xavier initialization of weight drastically reduces accuracy

5.) Feature Scaling lead to overflows