Vectorizing Logistic Regression's Gradient Descent Output

Vectoring Logistic Regression

To implement Logistic Regression, first we need to compute each derivative with respect to z:

$$dz^{(1)} = a^{(1)} - y^{(1)} \qquad dz^{(2)} = a^{(2)} - y^{(2)} \qquad \ldots \qquad dz^{(m)} = a^{(m)} - y^{(m)}$$

$$A = [a^{(1)} \quad a^{(2)} \quad \ldots \quad a^{(m)}] \text{ (matrix with the shape of (1 x m))}$$

$$Y = [y^{(1)} \quad y^{(2)} \quad \ldots \quad y^{(m)}] \text{ (matrix with the shape of (1 x m)}$$

$$dz = [dz^{(1)} \quad dz^{(2)} \quad \ldots \quad dz^{(m)}]$$

$$\Rightarrow dz = A - Y = [a^{(1)} - y^{(1)} \quad a^{(2)} - y^{(2)} \quad \ldots \quad a^{(m)} - y^{(m)}] \text{ (dz.shape = (1 x m))}$$

Recall:

$$dw = \frac{dL}{dz} = \frac{dL}{da} \cdot \frac{da}{dz} \cdot \frac{da}{dw} = dz \cdot x = (a - y) \cdot x$$

In the previous implementation, we need to loop over all of the training examples multiple times:

$$\begin{aligned} &dw = 0 \\ &dw \mathrel{+}= \ x_1 \cdot dz_1 \\ &dw \mathrel{+}= \ x_2 \cdot dz_2 \\ &\quad\vdots \\ &dw \mathrel{+}= \ x_m \cdot dz_m \\ &dw \mathrel{/}= \ m \end{aligned} \qquad\qquad \begin{aligned} &db = 0 \\ &db \mathrel{+}= \ dz_1 \\ &db \mathrel{+}= \ dz_2 \\ &\quad\vdots \\ &db \mathrel{+}= \ dz_m \\ &db \mathrel{/}= \ m \end{aligned}$$

Let's vectorize this:

$$\begin{aligned} db &= \tfrac{1}{m} \textstyle\sum_{i=1}^{m} \ dz^{(i)} \\ &= \tfrac{1}{m} \ np.sum(dz) \rightarrow \text{used to compute db} \end{aligned}$$

$$\begin{aligned} dw &= \tfrac{1}{m} \ X \cdot dz^T \rightarrow \text{used to compute dw} \\ &= \tfrac{1}{m} \cdot \begin{bmatrix} | & | & \ldots & | \\ x^{(1)} & x^{(2)} & \ldots & x^{(m)} \\ | & | & \ldots & | \end{bmatrix} \cdot \begin{bmatrix} dz^{(1)} \\ \cdot \\ \cdot \\ dz^{(m)} \end{bmatrix} \\ &= \tfrac{1}{m} \cdot [x^{(1)} dz^{(1)} \ + \ \ldots \ + \ x^{(m)} dz^{(m)}] \qquad \text{((n x 1) dimensional vector)} \end{aligned}$$

Implementing Logistic Regression:

non-vectorized method:

$$J = 0,\ dw_1 = 0,\ dw_2 = 0,\ db = 0$$

$$For\ i = 1\ \ to\ \ m:$$

$$z^{(i)} = w^T x^{(i)} + b$$

$$a^{(i)} = \sigma(z^{(i)})$$

$$J += \ -[y^{(i)} \log a^{(i)}\ +\ (1 - y^{(i)} \log(1 - a^{(i)})]$$

$$dz^{(i)} = a^{(i)} - y^{(i)}$$

$$dw_1 += x_1^{(i)} \cdot dz^{(i)}$$

$$dw_2 += x_2^{(i)} \cdot dz^{(i)}$$

$$(assuming\ that\ you\ have\ just\ 2\ features)$$

$$db += dz^{(i)}$$

$$J/ = m,\ \ dw_1/ = m,\ \ dw_2/ = m,\ \ db/ = m$$

Vectorized method:

$$z = w^T x + b$$

$$= np.dot(w.T,\ x)\ +\ b$$

$$A = \sigma(z)$$

$$dz = A - Y$$

$$dw = \frac{1}{m} X \cdot dz^T$$

$$db = np.sum(dz)$$

$$w = w\ -\ \alpha dw$$

$$b = b\ -\ \alpha db$$