Logistic Regression

Logistic Regression has form : $\theta(w^T x)$, where θ is an activation function

Sigmoid function

$$f(s) = \frac{1}{1+e^{-s}} \triangleq \sigma(s)$$

$$\lim_{s \to -\infty} \sigma(s) = 0 \qquad \lim_{s \to +\infty} \sigma(s) = 1$$

Specially, $$\sigma(s)' = \frac{e^{-s}}{(1+e^{-s})^2} = \frac{1}{1+e^{-s}} \cdot \frac{e^{-s}}{1+e^{-s}}$$

$$= \sigma(s)\left(1-\sigma(s)\right)$$

The value of sigmoid function is in range of $[-1, 1]$, but we can transform the range to $[0;1]$ by the function:

$$\tanh(s) = 2\sigma(2s) - 1$$

Loss function

Assume that the probability a data point x fall into class 1 is $f(w^T x)$, where f is a sigmoid function. Then, we have:

$$P(y_i = 1 \mid x_i ; w) = f(w^T x_i)$$

$$P(y_i = 0 \mid x_i ; w) = 1 - f(w^T x_i)$$

Let $z_i = f(w^T x_i)$, then :

$$P(y_i \mid x_i ; w) = z_i^{y_i}(1-z_i)^{1-y_i}$$

Suppose the data points are i.i.d, then:

$$P(Y \mid X ; w) = \prod_{i=1}^{N} P(y_i \mid x_i ; w)$$

$$= \prod_{i=1}^{N} z_i^{y_i}(1-z_i)^{1-y_i}$$

Our final problem becomes finding the minimum value of negative log likelihood function:

$$J(w) = -\log P(Y/X; w)$$
$$= - \sum_{i=1}^{n} \left(y_i . \log(z_i) + (1-y_i) \log(1-z_i) \right)$$

Optimize the loss function

The loss function of a data point i^{th} is:

$$J(y_i \mid x_i ; w) = -\left(y_i . \log(z_i) + (1-y_i) . \log(1-z_i)\right)$$

Take the derivative we have:

$$\frac{\partial J(y_i \mid x_i ; w)}{\partial w} = -\left(\frac{y_i}{z_i} + \frac{1-y_i}{1-z_i}\right) \frac{\partial z}{\partial w}$$

$$= \frac{z_i - y_i}{z_i(1-z_i)} . \frac{\partial z}{\partial w} \qquad (1)$$

Let $s = w^T x \Rightarrow z = f(s)$,

$$\frac{\partial z}{\partial w} = \frac{\partial z}{\partial s} . \frac{\partial s}{\partial w}$$

Because $\frac{\partial z}{\partial s} = z(1-z)$

$$\Leftrightarrow \frac{\partial z}{z(1-z)} = \partial s$$

$$\Leftrightarrow \frac{\partial z}{1-z} + \frac{\partial z}{z} = \partial s$$

$$\Leftrightarrow -\log(1-z) + \log(z) = s$$

$$\Leftrightarrow \log\left(\frac{z}{1-z}\right) = s$$

$$\Leftrightarrow \frac{z}{1-z} = e^s$$

$$\Leftrightarrow z = \frac{e^s}{1+e^s} = \frac{1}{1+e^{-s}} = \sigma(s)$$

$$(1) \Leftrightarrow \quad \frac{\partial J(y_i | x_i ; w)}{\partial w_i} = \frac{z_i - y_i}{z_i(1 - z_i)} \cdot \frac{\partial z}{\partial w}$$

$$= \frac{z_i - y_i}{z_i(1 - z_i)} \cdot \frac{\partial z}{\partial s} \cdot \frac{\partial s}{\partial w}$$

$$= (z_i - y_i)\, x_i$$

Finally, we could update w_i by SGD (stochastic gradient descent)

Vectorizing logistic regression

$X = [x_1, \dots, x_m]$ $\qquad$ $S = [s_1, \dots, s_m]$

$$S = w^T X + \overset{1 \times m}{[b, \dots, b]}$$

$$\rightarrow \quad S = \text{np.dot}(w.T, X) + b$$

$$Z = [z_1, \dots, z_m] = \sigma(S)$$

$$w -= (Z - Y)^T X$$

Implementation

for i : #iterations :

$$S = w^T X = \text{np.dot}(w.T, X)$$

$$Z = \sigma(S)$$

$$w -= \alpha \cdot X^T (Z - Y)$$