Voice recognition Algorithm

R Rohith Reddy

26 June 2019

$$y' = sigmoid(W.X + B)$$
$$y'i = sigmoid(\sum_i Wi * xi + b)$$

$$sigmoid(x) = 1/(1 + e^{-x})$$
$$sigmoidprime(x) = \frac{\partial sigmoid(x)}{\partial x}$$
$$= \frac{\partial}{\partial x}\frac{1}{1 + e^{-x}}$$
$$= \frac{e^{-x}}{(1 + e^{-x})^2}$$
$$= \frac{e^{-x}}{1 + e^{-x}}\frac{1}{1 + e^{-x}}$$
$$= \frac{e^{-x}}{1 + e^{-x}}\frac{1}{1 + e^{-x}}$$
$$= \frac{1 - 1 + e^{-x}}{1 + e^{-x}}\frac{1}{1 + e^{-x}}$$
$$= (1 - \frac{1}{1 + e^{-x}})\frac{1}{1 + e^{-x}}$$
$$= (1 - sigmoid(x))(sigmoid(x))$$

1 Gradient Descent method

$$J(W, b) = \frac{1}{2}\left\| y - \hat{y} \right\|^2$$

$$Wi := Wi - \alpha\frac{\partial J}{\partial W}$$
$$b := b - \alpha\frac{\partial J}{\partial b}$$

Partial differentiation terms are evaluated as follows:

$$\frac{\partial J}{\partial W} = \frac{\partial}{\partial W}\frac{1}{2}\left\| y - \hat{y} \right\|^2$$

$$\frac{\partial J}{\partial W} = \frac{\partial}{\partial W}\frac{1}{2}(y-\hat{y})^T(y-\hat{y})$$

$$\frac{\partial J}{\partial W} = \frac{1}{2}\frac{\partial}{\partial W}(\|y\|^2 - y^T\hat{y} - y\hat{y}^T + \|\hat{y}\|^2)$$

where

$$\frac{\partial \hat{y}}{\partial W} = \frac{\partial}{\partial W} sigmoid(W * x + b)$$

$$\frac{\partial \hat{y}}{\partial W} = sigmoidprime(W * x + b)x$$

$$\frac{\partial \hat{y}}{\partial b} = sigmoidprime(W * x + b)$$

Property:

$$\frac{\partial (xa)}{\partial x} = a$$

$$\frac{\partial (x^T x)}{\partial x} = 2x^T$$

$$= \frac{1}{2}(-\frac{\partial \hat{y}}{\partial W}y^T - \frac{\partial \hat{y}^T}{\partial W}y + \frac{\partial \hat{y}^T\hat{y}}{\partial W})$$

$$= \frac{1}{2}(-2\frac{\partial \hat{y}}{\partial W}y^T + 2\frac{\partial \hat{y}}{\partial W}\hat{y}^T)$$

$$= \frac{\partial \hat{y}}{\partial W}(-y^T + \hat{y}^T)$$

$$= sigmoidprime(W * x + b)x(-y + \hat{y})^T$$

$$= sigmoidprime(W * x + b)((-y + \hat{y})x^T)^T$$

$$= ((-y + \hat{y})x^T)^T sigmoidprime(W * x + b)$$

where

$$sigmoidprime(x) = sigmoid(x) * (1 - sigmoid(x))$$

Similarly for b matrix coefficients:

$$\frac{\partial J}{\partial b} = \frac{1}{2}\frac{\partial}{\partial b}(\|y\|^2 - y^T\hat{y} - y\hat{y}^T + \|\hat{y}\|^2)$$

$$= \frac{1}{2}(-\frac{\partial \hat{y}}{\partial b}y^T - \frac{\partial \hat{y}^T}{\partial b}y + \frac{\partial \hat{y}^T\hat{y}}{\partial b})$$

$$= \frac{1}{2}(-2\frac{\partial \hat{y}}{\partial b}y^T + 2\frac{\partial \hat{y}}{\partial b}\hat{y}^T)$$

$$= \frac{\partial \hat{y}}{\partial b}(-y^T + \hat{y}^T)$$

$$= sigmoidprime(W * x + b)(-y + \hat{y})^T$$

$$= (-y + \hat{y})^T sigmoidprime(W * x + b)$$

for these gradient descent method are:

$$W := W - \alpha((-y + \hat{y})x^T)^T sigmoidprime(W * x + b)$$

$$b := b - \alpha(-y + \hat{y})^T sigmoidprime(W * x + b)$$

$$W := W + \alpha((y - \hat{y})x^T)^T sigmoidprime(W * x + b)$$
$$b := b + \alpha(y - \hat{y})^T sigmoidprime(W * x + b)$$