CIS 5526 Joshua Schaaf

Discussed w/ Jared Meyers

Problem 1) Predictor: $\sum_{j=0}^{M} w_j x_j$

Loss: MSE $+ \alpha \sum_{j=0}^{M} w_j^2$

$$MSE = 1/N \sum_{i=1}^{M} (y_i - \hat{y}_i)^2$$
$$= 1/N \, E^T E$$
$$= 1/N (Y - \hat{Y})^T (Y - \hat{Y})$$

$$\vdots$$

$$MSE = 1/N (Y^T Y - 2w^T X^T Y + w^T X^T X w)$$

$$Loss = 1/N (Y^T Y - 2w^T X^T Y + w^T X^T X w) + \alpha \sum_{j=0}^{M} w_j^2$$

$$= 1/N (Y^T Y - 2w^T X^T Y + w^T X^T X w) + \alpha w^T w$$

this is in quadratic form, MSE is quadratic, and so is the reg. term, so their sum is quadratic.

$$\nabla Loss = \nabla MSE + \alpha \nabla w^T w$$

$$0 = -2/N \, X^T Y + 2/N \, X^T X w + 2 \alpha w$$
$$2/N \, X^T Y = 2/N \, X^T X w + 2 \alpha w$$
$$2/N \, X^T Y = (2/N \, X^T X + 2 \alpha I) w$$
$$1/N \, X^T Y = (1/N \, X^T X + \alpha I) w$$

$$\therefore w = (1/N \, X^T X + \alpha I)^{-1} \, 1/N \, X^T Y$$

CIS 5526: Homework 4

Problem 2) $relu(\sum_{j=0}^{M} \omega_j x_j)$ $relu(X\omega)$

M inputs, one neuron.

Because we are using relu as an activation function, MSE is piecewise

For $M = 1$ $relu(\sum_{j=0}^{M} \omega_j x_j) = relu(\omega x)$

$$MSE = \begin{cases} (y-\omega x)^2 = y^2 - 2y\omega x + \omega^2 x^2, & \omega x \geq 0 \\ y^2, & \omega x < 0 \end{cases}$$

Because MSE is piecewise, and y^2 is constant not quadratic, MSE is not in quadratic form.

For $\omega x \geq 0$: $\nabla MSE = -2yx + 2\omega x^2 = 0$

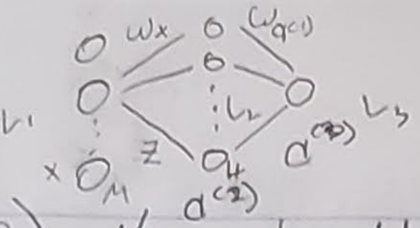
$$2\omega \cancel{x} = \cancel{2}y\cancel{x}$$

$$\omega = \frac{y}{x}$$

For $M = 1$ and
For $\omega x \geq 0$: $\omega \leftarrow \omega + \alpha(-2yx + 2\omega x^2)$

$MSE'' = 2x^2 \Rightarrow$ positive

$\therefore$ it is convex



$X w_x = Z_x$

Problem 3) N examples M features
H hidden relu nodes in one hidden layer,
and one sigmoid output neuron.

$$\text{Loss} = \sum_{i=1}^{N} y_i \log(nn(x_i;w)) + (1-y_i)\log(1-nn(x_i;w))$$

$$\frac{\partial \text{Loss}}{\partial w_2} = \left(\frac{\partial z^{(2)}}{\partial w_2}\right)\left(\frac{\partial a^{(3)}}{\partial z^{(2)}}\right)\left(\frac{\partial \text{Loss}}{\partial a^{(3)}}\right)$$

$$\frac{\partial z^{(2)}}{\partial w_2} = a_1 w_2 \quad \partial w_2 = a_1$$

$$\frac{\partial a^{(3)}}{\partial z^{(2)}} = \text{sigmoid}(z^{(2)})(1-\text{sigmoid}(z^{(2)}))$$

$$\frac{\partial \text{Loss}}{\partial a^{(3)}} = \frac{1}{N}\left(-\frac{y_i}{a^{(3)}} + \frac{1-y_i}{1-a^{(3)}}\right)$$

$$\frac{\partial \text{Loss}}{\partial w_1} = \left(\frac{\partial \text{Loss}}{\partial w_2}\right)\left(\frac{\partial z^{(1)}}{\partial w_1}\right)\left(\frac{\partial a^{(2)}}{\partial z^{(1)}}\right)\left(\frac{\partial \text{Loss}}{\partial a^{(2)}}\right)$$

$$\frac{\partial z^{(1)}}{\partial w_1} = x w_1 \quad \partial w_1 = X$$

$$\frac{\partial a^{(2)}}{\partial z^{(1)}} = \begin{cases} 0, & z^{(1)} < 0 \\ 1, & z^{(1)} \geq 0 \end{cases} = \text{relu}'(z^{(1)})$$

$$\frac{\partial \text{Loss}}{\partial w_1} = \frac{1}{N}\left(-\frac{y_i}{a^{(2)}} + \frac{1-y_i}{1-a^{(2)}}\right)$$