CSI 5130 - Homework 1

Adrian Sandoval

1/24/2020

Problem 1

Mean Squared Loss Function : $mse = L_2$

$L_2 = \frac{1}{N}\sum_{\forall_n}(w_0 + w_1x_n - t_n)^2$

Partial Div of: w_0

$\frac{\partial L_2}{\partial w_0} = \frac{1}{N}\sum_{\forall_n}(w_0 + w_1x_n - t_n)^2\frac{\partial L_2}{\partial w_0}$

$= \frac{1}{N} * \sum_{\forall_n}(w_0 + w_1x_n - t_n)^2\frac{\partial L_2}{\partial w_0}$

$= \frac{1}{N} * \sum_{\forall_n} 2(w_0 + w_1x_n - t_n) * (w_0 + w_1x_n - t_n)\frac{\partial L_2}{\partial w_0}$; **Chain Rule**

$= \frac{1}{N} * \sum_{\forall_n} 2(w_0 + w_1x_n - t_n) * (w_0\frac{\partial L_2}{\partial w_0} + w_1x_n\frac{\partial L_2}{\partial w_0} - t_n\frac{\partial L_2}{\partial w_0})$; **Chain Rule**

$= \frac{1}{N} * \sum_{\forall_n} 2(w_0 + w_1x_n - t_n) * (1 + 0 - 0)$

$= \frac{\sum_{\forall_n} 2(w_0+w_1x_n-t_n)}{N}$

If desired not to take the mean:

$= \sum_{\forall_n} 2(w_0 + w_1x_n - t_n)$

Solving For w_0

$0 = \frac{\sum_{\forall_n} 2(w_0+w_1x_n-t_n)}{N}$

$=> \sum_{\forall_n}(w_0 + w_1x_n - t_n)$

$0 = \sum_{\forall_n}(w_0 + w_1x_n - t_n)$

$\sum_{\forall_n} t_n - w_1x_n = w_0$

Partial Div of: w_1

$\frac{\partial L_2}{\partial w_1} = \frac{1}{N}\sum_{\forall_n}(w_0 + w_1x_n - t_n)^2\frac{\partial L_2}{\partial w_1}$

$= \frac{1}{N} * \sum_{\forall_n}(w_0 + w_1x_n - t_n)^2\frac{\partial L_2}{\partial w_1}$

$= \frac{1}{N} * \sum_{\forall_n} 2(w_0 + w_1x_n - t_n) * (w_0 + w_1x_n - t_n)\frac{\partial L_2}{\partial w_1}$; **Chain Rule**

$= \frac{1}{N} * \sum_{\forall_n} 2(w_0 + w_1x_n - t_n) * (w_0\frac{\partial L_2}{\partial w_1} + w_1x_n\frac{\partial L_2}{\partial w_1} - t_n\frac{\partial L_2}{\partial w_1})$; **Chain Rule**

$= \frac{1}{N} * \sum_{\forall_n} 2(w_0 + w_1x_n - t_n) * (0 + x_n - 0)$

$= \frac{\sum_{\forall_n} 2x_n(w_0+w_1x_n-t_n)}{N}$

If desired not to take the mean:

$= \sum_{\forall_n} 2x_n(w_0 + w_1x_n - t_n)$

Solving For w_1

$0 = \frac{\sum_{\forall_n} 2x_n(w_0+w_1x_n-t_n)}{N}$

$=> \sum_{\forall_n}(2x_nw_0 + 2x_nw_1x_n - 2x_nt_n)$

$0 = \sum_{\forall_n}(2x_nw_0 + 2x_nw_1x_n - 2x_nt_n)$

$\sum_{\forall_n} \frac{t_n-w_0}{x_n} = w_1$

Problem 2

$S = \{(1,1),(2,2),(3,3)\}$; $w_0 = 0,\ w_1 = 0$; $step = 0.1$

$w = w - \alpha(\frac{1}{2m}\sum_{(x,y)\varepsilon S}(y'-y)^2)$

Partial Derivative of w0

$\frac{\partial}{\partial w_0}J(w_0,w_1) = (\frac{1}{2m}\sum_{(x,y)\varepsilon S}(y'-y)^2)\frac{\partial}{\partial w_0}$

$= (\frac{1}{2m}\sum_{(x,y)\varepsilon S}(y'-y)^2)\frac{\partial}{\partial w_0} \quad => \frac{1}{2m}\sum_{(x,y)\varepsilon S}2(y'-y)*(y'-y)\frac{\partial}{\partial w_0}$

$= \frac{1}{2m}\sum_{(x,y)\varepsilon S}2(y'-y))*(w_0\frac{\partial}{\partial w_0} + w_1x_n\frac{\partial}{\partial w_0} - y\frac{\partial}{\partial w_0})$

$= \frac{1}{m}\sum_{(x,y)\varepsilon S}(y'-y))*(1+0-0) => \frac{1}{m}\sum_{(x,y)\varepsilon S}(y'-y)$

Partial Derivative of w1

$\frac{\partial}{\partial w_1}J(w_0,w_1) = (\frac{1}{2m}\sum_{(x,y)\varepsilon S}(y'-y)^2)\frac{\partial}{\partial w_1}$

$= (\frac{1}{2m}\sum_{(x,y)\varepsilon S}(y'-y)^2)\frac{\partial}{\partial w_1} => \frac{1}{2m}\sum_{(x,y)\varepsilon S}2(y'-y)*(y'-y)\frac{\partial}{\partial w_1}$

$= \frac{1}{2m}\sum_{(x,y)\varepsilon S}2(y'-y))*(w_0\frac{\partial}{\partial w_1} + w_1x_n\frac{\partial}{\partial w_1} - y\frac{\partial}{\partial w_1})$

$= \frac{1}{m}\sum_{(x,y)\varepsilon S}(y'-y))*(0+x_n-0) => \frac{1}{m}\sum_{(x,y)\varepsilon S}(y'-y)(x_n)$

Gradient Descent

3 Steps For w_0:

1. $w_0 = 0 - (0.1)\ (\frac{1}{3}((0-1)+(0-2)+(0-3))) = 0 - (-.2) = .2$
2. $w_0 = .2 - (0.1)\ (\frac{1}{3}((0.2-1)+(0.2-2)+(0.2-3))) = 0.2 - (-.18) = .38$
3. $w_0 = .38 - (0.1)\ (\frac{1}{3}((.38-1)+(.38-2)+(.38-3))) = .38 - (-.162) = .542$

3 Steps For w_1:

1. $w_1 = 0 - (0.1)(\frac{1}{3}([(0-1)*(1)+(0-2)*(2)+(0-3)*(3)])) = 0 - (-.467) = .467$
2. $w_1 = 0.467 - (0.1)(\frac{1}{3}([(.467-1)*(1)+(.467-2)*(2)+(.467-3)*(3)])) = 0.467 - (-.373) = 0.84$
3. $w_1 = 0.84 - (0.1)(\frac{1}{3}([(0.84-1)*(1)+(0.84-2)*(2)+(0.84-3)*(3)])) = .84 - (-.298) = 1.138$

So $y = 0.542 + 1.138x_n$ for 3 iterations

Stochastic Gradient Descent

For every point in the dataset S:

$w = w - \alpha(y'-y)x$

Steps for w_0

1. $w_0 = 0 - (0.1)(0 - 1) * 1 = 0.1$

2. $w_0 = 0.1 - (0.1)(0 - 2) * 2 = .5$

3. $w_0 = 0.5 - (0.1)(0 - 3) * 3 = 1.4$

$w_0 = 1.4$

Steps for w_1

1. $w_1 = 0 - (0.1)(0 - 1) * 1 = 0.1$

2. $w_1 = 0.1 - (0.1)(.1(2) - 2) * 2 = .46$

3. $w_1 = 0.46 - (0.1)(.46(3) - 3) * 3 = 0.946$

$w_1 = 0.946$