W3 tutorial

Univariate Linear Regression

Recall the formal statement of *univariate linear regression*:

- Given a training set $\{(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})\}$, train weights w_0, w_1 that minimise a loss function.
- Given this training set, and weights w_0, w_1, the *square loss* (or L_2 loss) function is given as

$$g(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}(w_0 + w_1 x^{(i)} - y^{(i)})^2.$$

- Informally, we need w_0, w_1 such that for all $i = 1, \ldots, n$

$$w_0 + w_1 x^{(i)} \approx y^{(i)}.$$

Multivariate Linear Regression

Recall the formal statement of *multivariate linear regression*:

- Given a training set $\{(\mathbf{x}^{(1)}, y^{(1)}), \ldots, (\mathbf{x}^{(n)}, y^{(n)})\}$, train a weight vector $\mathbf{w}$ that minimises a loss function.
- If we have d variables, then for all $i = 1, \ldots, n$, we write

$$\mathbf{x}^{(i)} = (1, x_1^{(i)}, x_2^{(i)}, \ldots, x_d^{(i)}) \text{ and } \mathbf{w} = (w_0, w_1, w_2, \ldots, w_d).$$

- Given this training set and a weight vector $\mathbf{w}$, the *square loss* (or L_2 loss) function is given as

$$g(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}^T\mathbf{x}^{(i)} - y^{(i)})^2.$$

- Informally, we need $\mathbf{w}$ such that for all $i = 1, \ldots, n$

$$\mathbf{w}^T\mathbf{x}^{(i)} \approx y^{(i)}.$$

Exercise 1

Consider the following pairs of points in the form (x, y). In each case, find the equation of the line that passes between the two given points in the form $y = ax + b$. Also, find its slope.

1. $(1, 2)$ and $(-1, -4)$.
2. $(-1, 3)$ and $(3, -5)$.
3. $(-2, -3)$ and $(1, 0)$.
4. $(3, 5)$ and $(0, 5)$.

Hint: You should find the values of a and b. The slope equals a.

1)
$a = \frac{2--4}{1--1} = \frac{6}{2} = 3$
$2 = (3 * 1) + b \Rightarrow b = -1$
$\therefore y = 3x - 1$
2)
$a = \frac{3--5}{-1-3} = \frac{8}{-4} = -2$
$3 = (-2 * -1) + b \Rightarrow b = 1$
$\therefore y = -2x + 1$
3)
$a = \frac{-3-0}{-2-1} = \frac{-3}{-3} = 1$
$-3 = (1 * -2) + b \Rightarrow b = -1$
$\therefore y = x - 1$
4)
$a = \frac{5-5}{3-0} = \frac{0}{3} = 0$
$5 = (0 * 3) + b \Rightarrow b = 2$
$\therefore y = 2$

Exercise 2

Consider a univariate linear regression problem with the square loss:

$$g(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}(w_0 + w_1 x^{(i)} - y^{(i)})^2$$

- We have this training set of size $n = 4$:

i	$x^{(i)}$	$y^{(i)}$
1	1	3
2	0	2
3	2	5
4	-1	0

Weights w_0, w_1	Loss $g(w_0, w_1)$
$w_0 = 2,\ w_1 = 3$	?
$w_0 = 3,\ w_1 = 1$	?
$w_0 = 2,\ w_1 = 2$	?
$w_0 = 0,\ w_1 = 2$	?

- Fill in the table to the right for each choice of weights.
- Which of these weights yield the minimum loss?

1)

$$\frac{1}{n}\sum_{i=1}^{n=4}(2 + (3 * x^{(i)}) - y^{(i)})^2$$

$$\frac{(2+(3*1)-3)^2+(2+(3*0)-2)^2+(2+(3*2)-5)^2+(2+(3*-1)-0)^2}{4}$$

$$\Rightarrow \frac{4+0+9+1}{4} \Rightarrow \frac{14}{4} = \frac{7}{2}$$

2)

$$\frac{1}{n}\sum_{i=1}^{n=4}(3 + (1 * x^{(i)}) - y^{(i)})^2$$

$$\frac{(3+(1*1)-3)^2+(3+(1*0)-2)^2+(3+(1*2)-5)^2+(3+(1*-1)-0)^2}{4}$$

$$\Rightarrow \frac{1+1+0+4}{4} \Rightarrow \frac{6}{4} = \frac{3}{2}$$

3)

$$\frac{1}{n}\sum_{i=1}^{n=4}(2 + (2 * x^{(i)}) - y^{(i)})^2$$

$$\frac{(2+(2*1)-3)^2+(2+(2*0)-2)^2+(2+(2*2)-5)^2+(2+(2*-1)-0)^2}{4}$$

$$\Rightarrow \frac{1+0+1+0}{4} \Rightarrow \frac{2}{4} = \frac{1}{2}$$

4)

$$\frac{1}{n}\sum_{i=1}^{n=4}(0 + (2 * x^{(i)}) - y^{(i)})^2$$

$$\frac{(0+(2*1)-3)^2+(0+(2*0)-2)^2+(0+(2*2)-5)^2+(0+(2*-1)-0)^2}{4}$$

$$\Rightarrow \frac{1+4+1+4}{4} \Rightarrow \frac{10}{4} = \frac{5}{2}$$

Exercise 3

Consider the following algorithm.

Algorithm 1: Single iteration of Gradient Descent for Univariate Linear Regression.

Input: Training set: $\{(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})\}$, learning rate: α

Output: Cost C; weights w_0, w_1.

$C \leftarrow 0$;
$w_0 \leftarrow 0$;
$w_1 \leftarrow 0$;
for $i = 1, \ldots, n$ **do**
  $f \leftarrow w_0 + w_1 x^{(i)}$;
  $C \leftarrow C + (f - y^{(i)})^2$;
  $w_0 \leftarrow w_0 - \alpha \cdot (f - y^{(i)})$;
  $w_1 \leftarrow w_1 - \alpha \cdot (f - y^{(i)}) x^{(i)}$.
return C, w_0, w_1.

- What are the numerical values of C, w_0, w_1 at the end of algorithm 1 for $\alpha = 1$ and the following training set of size $n = 3$:

i	$x^{(i)}$	$y^{(i)}$
1	1	1
2	2	5
3	3	11

Threshold is 0

1^{st} ***iteration:***

$f(1) = 0 + (0 * 1) = 0$

$C_{new} \leftarrow 0 + (0 - 1)^2 = 1$

$w_{0_{new}} \leftarrow 0 - 1 * (0 - 1) = 1$

$w_{1_{new}} \leftarrow 0 - 1 * (0 - 1) * 1 = 1$

2^{nd} ***iteration:***

$f(2) = 1 + (1 * 2) = 3$

$C_{new} \leftarrow 1 + (3 - 5)^2 = 5$

$w_{0_{new}} \leftarrow 1 - 1 * (3 - 5) = 3$

$w_{1_{new}} \leftarrow 1 - 1 * (3 - 5) * 2 = 5$

3^{rd} ***iteration:***

$f(3) = 3 + (5 * 3) = 18$

$C_{new} \leftarrow 5 + (18 - 11)^2 = 54$

$w_{0_{new}} \leftarrow 3 - 1 * (18 - 11) = -4$

$w_{1_{new}} \leftarrow 5 - 1 * (18 - 11) * 3 = -16$

Exercise 4

- Let $(\mathbf{x}, y)$ be a data point and $\mathbf{w}$ be the weight vector to be optimised in a multivariate linear regression model with d variables. Assume that $\mathbf{x}$ and $\mathbf{w}$ are of the form[1]

$$\mathbf{x} = (x_0, x_1, \ldots, x_d) \text{ and } \mathbf{w} = (w_0, w_1, \ldots, w_d).$$

- Let g be a square loss function of the form

$$g(\mathbf{w}) = (\mathbf{w}^T\mathbf{x} - y)^2.$$

- Use the derivative rules to prove that

$$\nabla g(\mathbf{w}) = 2(\mathbf{w}^T\mathbf{x} - y)\mathbf{x}.$$

- Hint: Find each partial derivative separately, then factor.

[1] We usually take $x_0 = 1$, but we leave it as x_0 here.

Using the chain rule:

$g(w) = (w^T x - y)^2$ is in the form $g(f(x))$

with $f(x) = w^T x - y$ and $g(x) = f(x)^2$

Differentiating $g(x)$ first gives $2f(x)$

Then differentiating $f(x)$ partially:

Firstly with respect w_0:

Gives:

x_0

Then w_1:

Gives:

x_1

Then w_d:

Gives:

x_d

Generalising that it means with respect to any w_i :

Gives:

$$x_i$$

Which means we can say it just differentiates to:

$$x$$

where x is $(x_0, x_1, ...x_d)$ with one value being actually $x_0/x_1/x_2$ and the rest being 0