ASSIGNMENT-1

SHASHANK REDDY DASARI

1) Function Approximation by Hand

$$(x,y) = \{(1,1),(2,2),(3,2),(4,5)\}$$

$$\hat{y} = \theta_1 x + \theta_2$$

$$J(\theta) = \frac{1}{N}\sum_{i=1}^{N}(y^i - \hat{y}^i)^2$$

→ $\theta = (1,0)$

$$\hat{y} = 1\cdot x + 0 = 1x$$

x	$\hat{y}$	$r = y - \hat{y}$	r^2
1	1	$1-1=0$	0
2	2	$2-2=0$	0
3	3	$2-3=-1$	1
4	4	$5-4=1$	1
		Σ	2

$$MSE = \frac{1}{4}(2) = 0.5$$

→ 2. $\theta = (0.5, 1)$

$$\hat{y} = 0.5x + 1$$

x	$\hat{y}$	$r = y - \hat{y}$	r^2
1	$0.5(1)+1 = 1.5$	$1-1.5=0.5$	$(-0.5)^2 = 0.25$
2	$0.5(2)+1 = 2$	$2-2=0$	0
3	$0.5(3)+1 = 2.5$	$2-2.5=-0.5$	$(-0.5)^2 = 0.25$
4	$0.5(4)+1 = 3$	$5-3=2$	$(2)^2 = 4$
			$\Sigma = 4.5$

$$MSE = \frac{1}{4}(4.5) = 1.125$$

→ 3. lower MSE is best fit

→ $\theta = (1,0)$ has best fit.

2. Random Guessing

$J(\theta_1, \theta_2) = 8(\theta_1 - 0.3)^2 + 4(\theta_2 - 0.7)^2$

$\Rightarrow$ 1. $J(0.1, 0.2)$ and $J(0.5, 0.9)$

$$J(0.1, 0.2) = 8(0.1-0.3)^2 + 4(0.2-0.7)^2$$
$$= 8(-0.2)^2 + 4(-0.5)^2$$
$$= 0.32 + 1$$
$$= 1.32$$

$$J(0.5, 0.9) = 8(0.5-0.3)^2 + 4(0.9-0.7)^2$$
$$= 8(0.2)^2 + 4(0.2)^2$$
$$= 0.32 + 0.16$$
$$= 0.48$$

$\rightarrow$ $(0.5, 0.9)$ is closer to the minimum $(0.3, 0.7)$

3) first Gradient Descent

Dataset: (1,3), (2,4), (3,6), (4,5)

$$\frac{\partial J}{\partial \theta_1} = -\frac{2}{N}\sum_{i=1}^{N} x^{(i)}(y^{(i)} - \hat{y}^{(i)}), \quad \frac{\partial J}{\partial \theta_2} = -\frac{2}{N}\sum_{i=1}^{N}(y^{(i)} - \hat{y}^{(i)})$$

$\theta \leftarrow \theta - \alpha \nabla J$

$\theta^{(0)} = (0,0) \quad \alpha = 0.01$

$\theta^0 = (0,0)$

$\theta^0 \quad \theta_1 = 0, \ \theta_2 = 0 \rightarrow \hat{y} = 0 \cdot x + 0 = 0$

$\hat{y}^{(1)} = 0$

$y^{(2)} = 0$

$y^{(3)} = 0$

$y^{(4)} = 0$

x	y	$\hat{y}$	$r = y - \hat{y}$	r^2
1	3	0	3	9
2	4	0	4	16
3	6	0	6	36
4	5	0	5	25

$\Sigma r = 18$

$\Sigma x r = 3 + 8 + 18 + 20 = 49$

∇J

$$\frac{\partial J}{\partial \theta_1} = -\frac{2}{N}\sum x^i r^i = \frac{-2}{4} \cdot 49 = -24.5$$

$$\frac{\partial J}{\partial \theta_2} = -\frac{2}{N}\sum r^i = \frac{-2}{4} \cdot 18 = -9$$

$\nabla J = (-24.5, -9)$

$\theta^{(1)} = \theta^0 - \alpha \nabla J = (0,0) - 0.01 \cdot (-24.5, -9) = (0.245, 0.09)$

$\Rightarrow \theta^1 = (0.245, 0.09)$

$J(\theta^{(0)})$

$\hat{y} = 0$

$$J = \frac{1}{4} \sum (y-0)^2 = \frac{1}{4}(3^2 + 4^2 + 6^2 + 5^2)$$

$$= \frac{1}{4}(86) = 21.5$$

$J(\theta^1)$

$\theta_1 = 0.245,\ \theta_2 = 0.09 \quad \hat{y} = 0.245x + 0.09$

$y_1 = 0.245(1) + 0.09 = 0.335$

$y_2 = 0.245(2) + 0.09 = 0.59$

$y_3 = 0.245(3) + 0.09 = 0.825$

$y_4 = 0.245(4) + 0.09 = 1.07$

$y - y_1 = \quad (3 - 0.335)^2 = (2.665)^2 = 7.10$

$(4 - 0.59)^2 = (3.41)^2 = 11.60$

$(6 - 0.825)^2 = (5.175)^2 = 26.78$

$(5 - 1.07)^2 = (3.93)^2 = 15.44$

$\overline{\quad 60.92 \quad}$

$$J(\theta^1) = \frac{60.92}{4} = 15.23$$

$J(\theta^0) = 21.5, \quad J(\theta^1) = 15.23.$

$\theta^1 = (0.245, 0.09)$

θ^1

$\hat{y}^1 = 0.335$

$\hat{y}^2 = 0.59$

$\hat{y}^3 = 0.825$

$\hat{y}^4 = 1.07$

$r_i = y^i - \hat{y}^i$

$r_1 = 3 - 0.335 = 2.665$

$r_2 = 4 - 0.59 = 3.41$

$r_3 = 6 - 0.825 = 5.175$

$r_4 = 5 - 1.07 = 3.93$

$\Sigma r = 15.18$

$\Sigma x r = 2.665 + 6.82 + 15.525 + 15.72 = 40.73$

∇J

$$\frac{\partial J}{\partial \theta_1} = -\frac{2}{4} \cdot 40.73 = -20.365$$

$$\frac{\partial J}{\partial \theta_2} = \frac{-2}{4} \cdot 15.18 = -7.59$$

$$\nabla J = (-20.366, -7.59)$$

θ^2

$$\theta^2 = \theta^1 - \alpha \nabla J = (0.245, 0.09) - 0.01(-20.365, -7.59)$$

$$= (0.245 + 0.20365, 0.09 + 0.0759)$$

$$= (0.44865, 0.1659)$$

$$\theta^2 = (0.4487, 0.1659)$$

$J(\theta^{(1)}), J(\theta^{(2)})$

$\rightarrow J(\theta^{(1)}) = 15.23$

$J(\theta^{(2)})$

$\hat{y} = 0.44865x + 0.1659$

$\hat{y}_1 = 0.44865(1) + 0.1659 = 0.61455$

$\hat{y}_2 = 0.44865(2) + 0.1659 = 1.0632$

$\hat{y}_3 = 0.44865(3) + 0.1659 = 1.51186$

$\hat{y}_4 = 0.44865(4) + 0.1659 = 1.9605$

3.

$r_1 = 3 - 0.61455 = 2.38545 \Rightarrow 5.690$

$r_2 = 4 - 1.0632 = 2.9368 \Rightarrow 8.626$

$r_3 = 6 - 1.51185 = 4.48815 \Rightarrow 20.143$

$r_4 = 5 - 1.9605 = 3.0395 \Rightarrow 9.238$

r^2 total: 43.697

$$J(\theta^{(2)}) = \frac{43.697}{4} = 10.924$$

$J(\theta^{(1)}) = 15.23$

$J(\theta^{(2)}) = 10.92$

4) Data = (1,2) (2,2) (3,4) (4,6)

MSE = $J(\theta) = \frac{1}{N} \sum_{i=1}^{N} (y^i - (\theta x^i + \theta_2))^2$

⇒ 1 $(\theta_1, \theta_2) = (0.2, 0.5)$ and $(\theta_1, \theta_2) = (0.9, 0.1)$

i. $(\theta_1, \theta_2) = (0.2, 0.5)$

$\hat{y} = \cancel{0.2x+0.5}\ 0.2x + 0.5$

x	$\hat{y}$	$r = y - \hat{y}$	r^2
1	$0.2(1)+0.5 = 0.7$	$2 - 0.7 = 1.3$	1.69
2	$0.2(2)+0.5 = 0.9$	$2 - 0.9 = 1.1$	1.21
3	$0.2(3)+0.5 = 1.1$	$4 - 1.1 = 2.9$	8.41
4	$0.2(4)+0.5 = 1.3$	$6 - 1.3 = 4.7$	22.09

$\Sigma r^2 = 33.40$

$J = \frac{1}{4}(33.40)$
$= 8.35$

ii $(\theta_1, \theta_2) = (0.9, 0.1)$

$\hat{y} = 0.9x + 0.1$

x	$\hat{y}$	$r = y - \hat{y}$	r^2
1	$0.9(1)+0.1 = 1$	$2 - 1 = 1$	$(1)^2 = 1$
2	$0.9(2)+0.1 = 1.9$	$2 - 1.9 = 0.1$	$(0.1)^2 = 0.01$
3	$0.9(3)+0.1 = 2.8$	$4 - 2.8 = 1.2$	$(1.2)^2 = 1.44$
4	$0.9(4)+0.1 = 3.7$	$6 - 3.7 = 2.3$	$(2.3)^2 = 5.29$

$\Sigma r^2 = 7.74$

$J = \frac{1}{4}(7.74) = 1.935$

$J(0.2, 0.5) = 8.35$

$J(0.9, 0.1) = 1.935 \approx 1.94$

2. Gradient Descent

$\theta = (0,0)$, $\alpha = 0.01$

$h = 0x + 0 = 0$

$i = (2, 2, 4, 6)$

$\sum x_i y_i = 1\cdot2 + 2\cdot2 + 3\cdot4 + 4\cdot6 = 42$

$\sum y_i = 2+2+4+6 = 14$

$$\frac{\partial J}{\partial \theta_1} = -\frac{2}{4}(42) = -21$$

$$\frac{\partial J}{\partial \theta_2} = -\frac{2}{4}(14) = -7 \qquad \nabla J\ (-21, -7)$$

$$\theta^1 = \theta^{(0)} - \alpha(-21, -7)$$

$$= (0,0) - 0.01(-21, -7)$$

$$\theta^1 = (0.21, 0.7)$$

$\hat{y} = 0.21x + 0.7$

x	$\hat{y}$	$r = y - \hat{y}$	r^2
1	$0.21(1)+0.7 = 0.28$	$2 - 0.28 = 1.72$	$(1.72)^2 = 2.9$
2	$0.21(2)+0.7 = 0.49$	$2 - 0.49 = 01.51$	$(1.51)^2 = 2.28$
3	$0.21(3)+0.7 = 0.70$	$4 - 0.70 = 3.30$	$(3.30)^2 = 10.89$
4	$0.21(4)+0.7 = 0.91$	$6 - 0.91 = 5.09$	$(5.09)^2 = 25.9$

$\sum r^2 = 42.03$

$$J(\theta^1) = \frac{1}{4}(42.03) = 10.50$$

→ $J(0.2, 0.5) = 8.35$

$J(0.9, 0.1) = 1.935 \approx 1.94$

from $(0,0)$ $\alpha = 0.01$

$J(0.21, 0.07) = 10.51$

3.

1. TD

The random guess (0.9, 0.1) gave the lowest error $J = 1.935$ which is much lower than other random guess and first step gradient descent step.

Gradient descent starts at (0,0) predicting zeros with small step ($\alpha = 0.01$), $\theta' = (0.21, 0.07)$ stays near zero and gradient descent is systematic but slow initally with small learning rate

While a good random guess can land near the optimum.

5. Recognizing Underfitting and overfitting

1. Underfitting

2. underfitting occurs when the model is too simple to capture the underlying patterns in data

→ It fails to fit the trainning data well, leading to high training error

→ consequently it also generalize poorly. so test error remains high

3. → use a more complex model

→ Relax Regularization if its too strong.

6. Comparing Models

1. Model A is overfitting, it fits training data perfectly but fails on new data

Model B: is Underfitting, performs poorly on both training and test data

2. Model A

Low bias, model is flexible enough to fit training data

High variance, model is sensitive to the specifics of training sample

Model B:

High Bias, model is too simple

low variance: (not sensitive to the specifics of training data)

3. Model A (Improvements)

→ Simplify the model

→ Strengthen Regularization

→ Increase Training data

Model B (Improvements)

→ Increase Model capacity

→ Use complex model

→ Add Informative features

→ Reduce Regularization.