Chapter. 4

Single-variate Linear Regression

Theory Assignment

FAST CAMPUS
ONLINE

강사. 신경식

Life Changing Education

Chapter. 4

Loss Functions

FAST CAMPUS
ONLINE
신경식 강사.

Question. 4−01

Dataset이 다음과 같이 주어졌다고 하자,

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)})\} = \{(1,3), (3,7)\}$$

위의 Dataset은 $y = 2x + 1$에서부터 만들어졌기 때문에, 모델을 $\hat{y} = \theta_1 x + \theta_0$로 설정하였다.

initial $\vec{\theta}$가 $\theta_1 = -1, \theta_0 = -1$일 때, 다음 질문에 답하시오.

1) 각 Data sample에 대해 Square error를 이용한 Loss를 구하시오.

2) Loss function이 θ_1, θ_0에 대해 각각에 대해 몇 차 식인지 구하고 Convexity를 확인하시오.

3) 각 Data sample에 대해 $\vec{\theta} = (\theta_1, \theta_0) = (1,0)$일 때와 $\vec{\theta} = (\theta_1, \theta_0) = (2,1)$일 때의 Loss를 각각 구하고 비교하시오.

4) θ_1이 +1 커진 경우와 θ_0가 +1 커진 경우 중 어느 경우가 Loss가 더 많이 감소하는지 비교하시오.

Question. 4-02

$x > 0$인 Dataset $\mathcal{D}$가 다음과 같이 주어졌다.

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)})\} = \{(0.1, 3.1), (1, 4), (5, 8)\}$$

Dataset을 $y = x + 3$에서부터 만들었기 때문에, 모델을 $\hat{y} = \theta_1 x + \theta_0$로 설정하였다.

initial $\vec{\theta}$가 $\theta_1 = -1, \theta_0 = -1$일 때, 다음 질문에 답하시오.

1) 각 data samples에 대한 Square Loss를 구하고 서로 비교하시오.

2) 1)에서의 결과를 통하여 $x > 0$인 Dataset에서 $|x^{(i)}|$가 γ배 되었을 때, Loss의 변화를 증가, 감소로 표현하시오.

Life Changing Education

Question. 4-03

$x < 0$인 Dataset $\mathcal{D}$가 다음과 같이 주어졌다.

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)})\} = \{(-0.1, 2.9), (-1, 2), (-5, -2)\}$$

Dataset을 $y = x + 3$에서부터 만들었기 때문에, 모델을 $\hat{y} = \theta_1 x + \theta_0$로 설정하였다.

initial $\vec{\theta}$가 $\theta_1 = -1, \theta_0 = -1$일 때, 다음 질문에 답하시오.

1) 각 data samples에 대한 Square Loss를 구하고 서로 비교하시오.

2) 1)에서의 결과를 통하여 $x < 0$인 Dataset에서 $|x^{(\mathrm{i})}|$가 γ 배 되었을 때, Loss의 변화를 증가, 감소로 표현하시오.

Chapter. 4

Gradient Descent

Question. 4−04

$|x| \geq 1$인 Dataset $\mathcal{D}$가 다음과 같이 주어졌다.

$$\mathcal{D} = \{(1,7),(5,15),(-1,3),(-5,-5)\}$$

Dataset을 $y = 2x + 5$에서부터 만들었기 때문에, 모델을 $\hat{y} = \theta_1 x + \theta_0$로 설정하였다.

initial $\vec{\theta}$=(θ_1, θ_0) =(1,1)이고, learning rate α =0.1로 주어졌을 때 다음 질문에 답하시오.

1) Square error를 Loss function으로 사용하였을 때, $\frac{\partial \mathcal{L}}{\partial \theta_1}$과 $\frac{\partial \mathcal{L}}{\partial \theta_0}$을 구하시오.

2) θ_1과 θ_0에 대한 gradient vector의 norm을 구하시오.

3) θ_1 ,θ_0의 Gradient Descent Method를 구하시오.

4) 각 Data sample에 의한 θ_1 ,θ_0의 Update 양을 구하시오.

5) 4)에서의 결과를 통해 $x^{(i)}$가 γ 배 되었을 때, θ_1, θ_0가 Update 되는 양의 변화를 구하시오.

Question. 4−05

$0 \leq |x| < 1$인 Dataset $\mathcal{D}$가 다음과 같이 주어졌다.

$$\mathcal{D} = \{(0, 5), (0.1, 5.2), (0.3, 5.6), (-0.1, 4.8), (-0.3, 4.4)\}$$

Dataset을 $y = 2x + 5$에서부터 만들었기 때문에, 모델을 $\hat{y} = \theta_1 x + \theta_0$로 설정하였다.

initial $\vec{\theta}=(\theta_1, \theta_0) = (1,1)$이고, learning rate $\alpha = 0.1$로 주어졌을 때 다음 질문에 답하시오.

1) 각 Data sample에 의한 θ_1 ,θ_0의 Update 양을 구하시오.

2) 1)에서의 결과를 통해 $x^{(i)}$가 γ 배 되었을 때, θ_1, θ_0가 Update 되는 양의 변화를 구하시오.

Question. 4−06

Question. 4−04와 Question. 4−05의 결과를 토대로 $|x| \geq 1$인 경우와 $0 \leq |x| < 1$ 인 경우에서 θ_1과 θ_0 중 어떤 learnable parameter의 학습이 더 주도적으로 일어나는지 설명하시오.

또한, 이를 토대로 $|x| \geq 1$인 경우와 $0 \leq |x| < 1$ 인 경우에서 각각 θ_1, θ_0 중 어떤 것이 더 발산할 위험성이 높은지 설명하시오.

Question. 4−07

target function $y = \theta_1^* x + \theta_0^*$의 Linear Regression을 위해

Prediction 모델을 $\hat{y} = \theta_1 x + \theta_0$로, Loss를 Square Error로 설정하였다.

이때, 다음 질문에 답하시오.

1) $\underset{(\theta_1, \theta_0)}{argmin} \mathcal{L}$를 만족시키는 θ_1, θ_0이 θ_1^*, θ_0^*임을 보이시오.

2) initial $\vec{\theta} = (\theta_1, \theta_0)$가 (−1,−1)로 주어졌을 때, 다음과 같은 target function에서도 학습이 잘 이루어짐을 보이시오.

이때, learning rate = 0.1, data sample은 $x = 1$인 경우에 대하여 1 iteration동안 학습을 진행하시오.

$y = -x + 1$

$y = 3x - 1$

$y = 3x + 5$

Question. 4−08

target function $y = 3x + 2$의 이상적인 Dataset이 다음과 같다.

$$\mathcal{D} = \{(1,5),(2,8),(5,17)\}$$

그러나 noise에 의해 Dataset이 다음과 같이 왜곡된 경우를 가정하자.

$$\mathcal{D} = \{(1,6),(2,5),(5,15)\}$$

θ_1, θ_0이 (3,2)로 학습이 끝난 상태에서 각 data sample에 의해 θ_1, θ_0가 update되는 값들을 구하고,

bias가 있는 경우에서 noise가 학습에 미치는 영향을 분석하시오.

Chapter. 4

Learning with One Sample

Question. 4−09

2개의 Dataset $\mathcal{D}_1$과 $\mathcal{D}_2$가 다음과 같이 주어졌다.

$$\mathcal{D}_1 = \{(-0.5, 2.5), (0.5, 3.5)\}$$

$$\mathcal{D}_2 = \{(-1, 2), (1, 4)\}$$

$\mathcal{D}_1, \mathcal{D}_2$ 모두 $y = x + 3$에서부터 만들었기 때문에, 모델을 $\hat{y} = \theta_1 x + \theta_0$로 설정하였다.

initial $\vec{\theta} = (\theta_1, \theta_0) = (-1, -1)$이고, learning rate $\alpha = 0.1$로 주어졌을 때 다음 질문에 답하시오.

1) Loss에 대한 Update Equation을 이용하여 1번의 epoch동안 $\mathcal{D}_1$, $\mathcal{D}_2$ 각각 $\vec{\theta}$의 변화를 구하시오.

2) $\mathcal{D}_1$과 $\mathcal{D}_2$중 어느 Dataset이 $\vec{\theta}$를 target $\vec{\theta}^*$에 더 가깝게 하였는지 판단하고 이유를 설명하시오.

Question. 4−10

2개의 Dataset $\mathcal{D}_1$과 $\mathcal{D}_2$가 다음과 같이 주어졌다.

$$\mathcal{D}_1 = \{(-3,0),(3,6)\}$$

$$\mathcal{D}_2 = \{(-10,-7),(10,13)\}$$

Question. 4−09과 마찬가지로 $\mathcal{D}_1, \mathcal{D}_2$ 모두 $y = x + 3$에서부터 만들었기 때문에

모델을 $\hat{y} = \theta_1 x + \theta_0$로 설정하였다.

initial $\vec{\theta}$=(θ_1, θ_0) =(−1,−1)이고, learning rate α =0.1로 주어졌을 때 다음 질문에 답하시오.

1) Loss에 대한 Update Equation을 이용하여 1번의 epoch동안 $\mathcal{D}_1$, $\mathcal{D}_2$ 각각 $\vec{\theta}$의 변화를 구하시오.

2) Q.4−9와 Q.4−10의 결과를 토대로 Q.4−10의 $\mathcal{D}_1, \mathcal{D}_2$에서 $\vec{\theta}$가 발산하는 이유를 설명하시오.

Chapter. 4

Cost Functions

Question. 4−11

Linear regression을 위한 Dataset이 다음과 같이 주어졌다.

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(n)}, y^{(n)})\}$$

이때, Dataset은 $y = ax + b$에서부터 만들어졌다. 따라서 Linear regression을 위한 Prediction 모델은 $\hat{y} = \theta_1 x + \theta_0$, Loss는 Square Error, Cost는 MSE를 사용할 수 있다.

$\vec{\theta}$를 Update하기 위해 n개의 data sample을 이용할 때, 1번의 iteration동안 $\vec{\theta}$가 dataset을 잘 표현하는 $\vec{\theta}$로 Update되는 과정을 설명하시오.

단, forward/backward propagation을 설명하기 위해 각 연산은 basic building node를 이용하시오.