Chapter. 3

Single-variate Linear Regression without Bias Term

Theory Assignment

FAST CAMPUS ONLINE

강사. 신경식

Question.3-01

Dataset이 다음과 같이 단 하나의 data sample로만 이루어졌다고 하자.

$$\mathscr{D} = \{(x^{(1)}, y^{(1)})\} = \{(1,2)\}$$

위의 dataset은 $y = 2x$에서부터 만들어졌기 때문에, model을 $\hat{y} = \theta x$로 설정할 때 다음 문제들에 답하시오.

1) θ가 1, 1.5, 2일 때의 square error를 이용한 loss를 각각 구하고, θ가 2에 가까워질때 loss의 변화를 비교하시오.

(square error는 y와 $\hat{y}$를 이용하여 $(y - \hat{y})^2$로 구한다.)

2) 임의의 θ에 대한 loss를 algebraic equation으로 표현하고 그래프를 그리시오.

Question.3-02

Dataset이 다음과 같이 주어졌다고 하자.

$$\mathscr{D} = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)})\} = \{(1,2), (2,4), (3,6)\}$$

위의 dataset에 대해 다음의 물음에 답하시오.

1) model을 $\hat{y} = x$로 설정한다면, 각 data sample들에 대한 square loss를 구하고 서로 비교하시오.

2) square error loss를 임의의 data point $(x^{(i)}, y^{(i)})$에 대한 algebraic equation으로 표현하고,

$(x^{(i)}, y^{(i)})$에 대한 square error loss를 최소로 만드는 θ를 구하시오.

단, θ는 변수로 사용하여 model은 $\hat{y} = \theta x$로 설정하시오.

3) 2)에서 구한 algebraic equation이 θ에 대한 몇차인지 구하고, convexity를 말하시오.

Question.3-03

문제에서의 상황이 Question.3-02와 동일할 때, 다음의 물음에 답하시오.

1) Question.3-02의 1) 상황에서 1)의 결과를 이용하여 MSE cost를 구하시오.

2) MSE cost를 algebraic equation으로 표현하고, MSE cost를 최소로 만드는 θ를 구하시오.

3) 2)에서 구한 algebraic equation이 θ에 대한 몇차인지 구하고, convexity를 말하시오.

Question.3-04

다음과 같이 하나의 data sample만 가지고 있는 Dataset이 주어졌다.

$$\mathscr{D} = \{(1,2)\}$$

이때 다음 질문들에 답하시오. 단 prediction model은 $\hat{y} = \theta x$를 사용한다.

1) Square error를 loss로 사용할 때 loss에 대한 식을 구하고, $\dfrac{\partial \mathscr{L}(\theta)}{\partial \theta}$의 식을 구하시오.

   그리고 gradient descent method를 이용하여 θ를 update시키는 식을 구하시오.

2) initial θ를 1로, learning rate을 0.1로 설정했을 때 3 iteration 동안 update되는 θ들을 구하시오.

   그리고 target θ에 가까워지는지 확인하시오.

3) loss의 식을 graph로 표현하고 2)에서 구한 θ가 update되는 위치들을 이 graph 위에 나타내시오.

Question.3-05

다음과 같이 하나의 data sample만 가지고 있는 Dataset이 주어졌다.

$$\mathscr{D} = \{(1,2)\}$$

이때 다음 질문들에 답하시오.

단 Question.3-04와 마찬가지로 prediction model은 $\hat{y} = \theta x$를 사용하고, initial θ는 1로 설정한다.

1) learning rate이 0.8일때, 3 iteration 동안 update되는 θ들을 구하시오.

2) learning rate이 1.1일때, 3 iteration 동안 update되는 θ들을 구하시오.

Question.3-06

다음과 같이 Dataset이 주어졌다.

$$\mathscr{D} = \{(1,2), (3,6), (4,8)\}$$

이때 다음 질문들에 답하시오.

단 Question.3-04, Question.3-05와 마찬가지로 prediction model은 $\hat{y} = \theta x$를 사용하고, initial θ는 1로 설정한다.

1) learning rate이 0.1일때, 각각 θ의 loss에 대한 update equation을 구하시오.

2) 1)에 구한 update equation을 이용하여 3번의 iteration에 대해 각각 θ의 변화를 구하고, Question.3-05 3)의 관점에서 data sample에 따른 학습의 불안정성을 설명하시오.

3) learning rate이 0.01일 때, 각각 θ의 loss에 대한 update equation을 구하고, 3번의 iteration에 대해 각각 θ의 변화를 구하시오. 추가로 2)에서의 학습의 불안정성이 해결되는지 설명하시오.

Question.3-07

다음과 같이 Dataset이 주어졌다.

$$\mathscr{D} = \{(1,2), (3,6), (4,8)\}$$

이때 다음 질문들에 답하시오.

단 Question.3-04, Question.3-05와 마찬가지로 prediction model은 $\hat{y} = \theta x$를 사용하고, initial θ는 1로 설정한다.

1) 각 3개의 data point들의 loss를 이용하여 cost J를 구하고, $\dfrac{\partial J(\theta)}{\partial \theta}$를 구하시오.

2) $\dfrac{\partial J(\theta)}{\partial \theta}$와 gradient descent method를 이용한 θ의 update equation을 구하고,
$\alpha = 0.1$일때 4 iterations에 대한 θ의 변화를 구하시오.

Question.3-08

Question.3-08의 질문들은 Question.3-06과 Question.3-07을 바탕으로 해결하시오.

1) $\frac{\partial J(\theta)}{\partial \theta}$와 $\frac{\partial \mathcal{L}^{(1)}(\theta)}{\partial \theta}$, $\frac{\partial \mathcal{L}^{(2)}(\theta)}{\partial \theta}$, $\frac{\partial \mathcal{L}^{(3)}(\theta)}{\partial \theta}$의 관계를 설명하고, 이들을 이용한 gradient descent methods

$$\theta := \theta - \alpha \frac{\partial \mathcal{L}^{(i)}(\theta)}{\partial \theta} \qquad \theta := \theta - \alpha \frac{\partial J(\theta)}{\partial \theta}$$

의 관계를 설명하시오.

2) Question.3-06에서 (4,8)에 대한 loss를 이용하여 θ를 학습시키면 학습이 되지 않았다.

하지만 Question.3-07에서는 cost를 이용하여 θ를 학습시킬 때 (4,8)이 사용되었는데 올바르게 학습이 되었다.

두 과정의 차이점을 설명하시오.

Question.3-09

Linear regression을 위한 dataset이 다음과 같이 주어졌다.

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

이때, dataset은 $y = ax$에서부터 만들어졌다.

따라서 linear regression을 통해 predictor를 학습시킬때, model은 $\hat{y} = \theta x$, loss는 square error를 사용할 수 있다.

θ를 update하기 위해 하나의 data sample만 이용할때, 1번의 iteration에 대해 θ가 dataset을 잘 표현하는

θ로 update되는 과정을 설명하시오.

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

Question.3-10

Linear regression을 위한 dataset이 다음과 같이 주어졌다.

$$D = \{(1,3), (3,9), (2,6), (-1,-3)\}$$

learning rate을 0.01로, initial θ는 1로 설정하고, Question.3-09와 같은 방법으로 학습을 진행할 때

다음 질문들에 답하시오.

1) 1 epoch 동안 각 data sample들에 대해 loss와 θ가 update되는 절댓값을 구하시오.

그리고 loss의 감소 graph의 fluctuation이 생기는 이유를 설명하시오.

2) 1)의 결과를 통하여 $x^{(i)}$의 크기가 γ배 되었을 때, loss와 θ가 update되는 양은 몇 배가 되는지 구하시오.

Question.3-11

Linear regression을 통해 predictor를 학습시킬 때, iteration이 지날 때마다 loss의 감소량과

θ가 update되는 양이 줄어드는 이유를 설명하시오.

Question.3-12

대부분의 dataset에는 noise가 섞여있다. $y = 2x$에서 이상적으로 만들어진 Dataset이 다음과 같을 때

$$D = \{(1,2), (2,4), (3,6)\}$$

noise에 의해 실제로 수집한 dataset이 다음과 같이 왜곡되었다고 가정하자.

$$D = \{(1,0.5), (2,5), (3,5)\}$$

θ가 2이었을때, 즉 predictor가 target function일때 각 data sample에 의해 θ가 update되는 값들을 구하고,

noise가 학습에 미치는 영향을 분석하시오.

Question.3-13

Linear regression을 위한 dataset이 다음과 같이 주어졌다.

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

이때, dataset은 $y = ax$에서부터 만들어졌다.

따라서 linear regression을 통해 predictor를 학습시킬때,

model은 $\hat{y} = \theta x$, loss는 square error, cost는 MSE를 사용할 수 있다.

θ를 update하기 위해 2개의 data sample를 이용할때, 1번의 iteration에 대해 θ가 dataset을 잘 표현하는

θ로 update되는 과정을 설명하시오.

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

Question.3-14

Linear regression을 위한 dataset이 다음과 같이 주어졌고, 이 dataset을 이용하여 predictor를 학습시키려고 한다.

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

이때 loss를 사용하여 θ를 update시키면 $(x^{(i)}, y^{(i)})$의 크기에 따라 θ를 θ^*에서 멀어지게 하는 $(x^{(i)}, y^{(i)})$가 존재할 수 있다.

cost를 사용하게되면 이 문제에 대한 위험성을 낮출 수 있는데, 그 이유를 설명하시오.

Question.3-15

Linear regression을 위한 dataset이 다음과 같이 주어졌다.

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

Question.3-13와 같은 방법으로 학습을 진행할 때, n개의 data sample을 이용하여 θ를 update한다면

loss의 감소에서 fluctuation은 어떻게 변하는지 설명하시오.

Question.3-16

Linear regression을 위한 dataset이 다음과 같이 주어졌다.

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

이때, dataset은 $y = ax$에서부터 만들어졌다.

따라서 linear regression을 통해 predictor를 학습시킬때,

model은 $\hat{y} = \theta x$, loss는 square error, cost는 MSE를 사용할 수 있다.

θ를 update하기 위해 n개의 data sample를 이용할때, 1번의 iteration에 대해 θ가 dataset을 잘 표현하는

θ로 update되는 과정을 Vector notation을 이용하여 설명하시오.

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

Question.3-17

Question.3-18에서는 Jacobian들의 matrix multiplication을 이용하여 $y = \theta x$를 학습시키는 과정을 증명하였다.

그리고 이 과정에서 $\frac{\partial \overrightarrow{z_2}}{\partial \overrightarrow{z_1}}, \frac{\partial \overrightarrow{\mathscr{L}}}{\partial \overrightarrow{z_2}}$는 모두 diagonal matrix인 것을 확인할 수 있었다.

이 관점에서 다음의 질문들에 답하시오.

1) nxn diagonal matrix A, B의 matrix multiplication은 각 diagonal entry들을 원소로 가지는 vector들의 Hadamard product으로 표현할 수 있음을 보이시오.

2) Question.3-16의 과정을 1)의 내용을 이용하여 Hadamard product으로 표현하시오.