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.3-10 1)의 결과에서 data sample $(x^{(i)}, y^{(i)})$에 대해

$$\mathcal{L}^{(i)} = (y^{(i)} - \theta \cdot x^{(i)})^2$$

$$\theta := \theta + 2\alpha x^{(i)}(y^{(i)} - \theta \cdot x^{(i)})$$

의 식을 통해 loss가 구해지고 θ가 update되기 때문에 loss의 감소에서 fluctuation이 심하게 일어났다.

한편 위 dataset에서 n개의 data sample을 이용해 cost를 구하고 θ를 update시키면

$$J = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - \theta \cdot x^{(i)})^2$$

$$\theta := \theta + \frac{\alpha}{n}\sum_{i=1}^{n} 2x^{(i)}(y^{(i)} - \theta \cdot x^{(i)})$$

가 된다. 그리고 J의 경우 n개의 loss들의 평균을 구하고 θ의 update 경우에도 $\frac{\partial \mathcal{L}^{(i)}}{\partial \theta}$를 평균적으로 반영하기 때문에 n이 커질수록 loss가 크거나 작은 개별 data sample을 반영하지 않고 dataset의 전반적인 특성을 따르게 된다.
따라서 n이 커질수록 $(y^{(i)} - \theta \cdot x^{(i)})^2$와 $2x^{(i)}(y^{(i)} - \theta \cdot x^{(i)})$들이 평균화되므로 fluctuation이 점점 사라진다.
그러다 n이 dataset 전체 data sample의 개수가 되면 fluctuation은 없어지고 cost는 부드럽게 감소하게 된다.