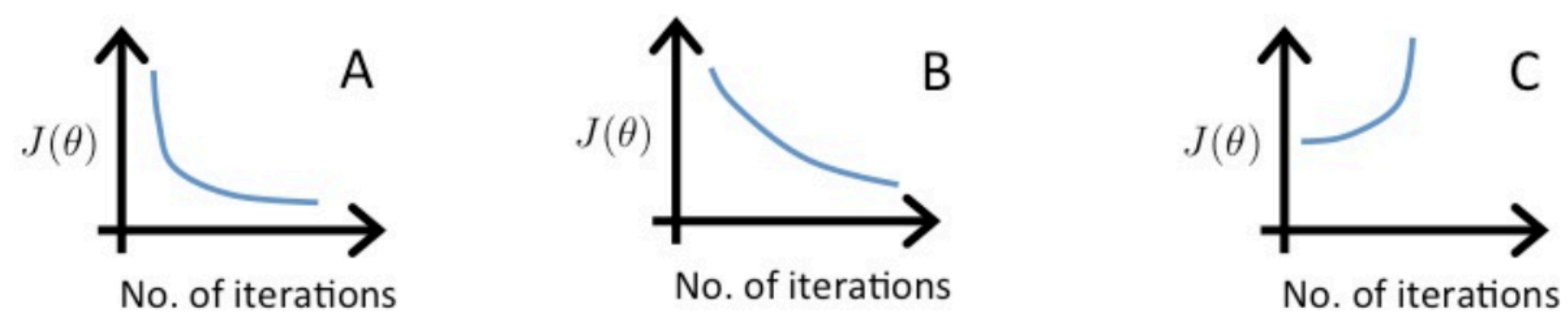
Suppose a friend ran gradient descent three times, with $\alpha = 0.01$, $\alpha = 0.1$, and $\alpha = 1$, and got the following three plots (labeled A, B, and C):



Which plots corresponds to which values of α?

- A is $\alpha = 0.01$, B is $\alpha = 0.1$, C is $\alpha = 1$.
- A is $\alpha = 0.1$, B is $\alpha = 0.01$, C is $\alpha = 1$.

Correct

In graph C, the cost function is increasing, so the learning rate is set too high. Both graphs A and B converge to an optimum of the cost function, but graph B does so very slowly, so its learning rate is set too low. Graph A lies between the two.

- A is $\alpha = 1$, B is $\alpha = 0.01$, C is $\alpha = 0.1$.
- A is $\alpha = 1$, B is $\alpha = 0.1$, C is $\alpha = 0.01$.

Size (feet)^2	Number of bedrooms	Number of floors	Age of home (years)	Price ($1000)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
...	...	...	...	...

In the training set above, what is $x_1^{(4)}$?

- The size (in feet^2) of the 1st home in the training set
- The age (in years) of the 1st home in the training set
- The size (in feet^2) of the 4th home in the training set

Correct

Continue

When there are n features, we define the cost function as

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2.$$

For linear regression, which of the following are also equivalent and correct definitions of $J(\theta)$?

- [x] $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (\theta^T x^{(i)} - y^{(i)})^2$

Correct

- [x] $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(\left(\sum_{j=0}^{n} \theta_j x_j^{(i)} \right) - y^{(i)} \right)^2$ (Inner sum starts at 0)

Correct

- [ ] $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(\left(\sum_{j=1}^{n} \theta_j x_j^{(i)} \right) - y^{(i)} \right)^2$ (Inner sum starts at 1)

Un-selected is correct

- [ ] $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(\left(\sum_{j=0}^{n} \theta_j x_j^{(i)} \right) - \left(\sum_{j=0}^{n} y_j^{(i)} \right) \right)^2$

Un-selected is correct

Suppose you are using a learning algorithm to estimate the price of houses in a city. You want one of your features x_i to capture the age of the house. In your training set, all of your houses have an age between 30 and 50 years, with an average age of 38 years. Which of the following would you use as features, assuming you use feature scaling and mean normalization?

- ○ $x_i = \text{age of house}$
- ○ $x_i = \dfrac{\text{age of house}}{50}$
- ○ $x_i = \dfrac{\text{age of house} - 38}{50}$
- ◉ $x_i = \dfrac{\text{age of house} - 38}{20}$

Correct

Suppose you want to predict a house's price as a function of its size. Your model is

$h_\theta(x) = \theta_0 + \theta_1(\text{size}) + \theta_2\sqrt{(\text{size})}$.

Suppose size ranges from 1 to 1000 (feet^2). You will implement this by fitting a model

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$.

Finally, suppose you want to use feature scaling (without mean normalization).

Which of the following choices for x_1 and x_2 should you use? (Note: $\sqrt{1000} \approx 32$.)

- $x_1 = \text{size},\ x_2 = 32\sqrt{(\text{size})}$
- $x_1 = 32(\text{size}),\ x_2 = \sqrt{(\text{size})}$
- $x_1 = \frac{\text{size}}{1000},\ x_2 = \frac{\sqrt{(\text{size})}}{32}$

Correct

- $x_1 = \frac{\text{size}}{32},\ x_2 = \sqrt{(\text{size})}$.