Problem 1

1, Cost function:

2,

Problem 2

1. Feature Normalization
2. Mean:

Standard Deviation:

1. Please refer to python code.
2. Please refer to python code.
3. Gradient Descent to Find Weights (Parameters)
4. with
5. Please refer to python code.

With

With

With

With

With

When , the convergence rate is the best among all five different learning rates, and the cost at 70 iterations is the best among all five learning rates.

1. Predicting Housing Prices

Using and 70 iterations, the prediction price is 29592277.531

1. Stochastic Gradient

Please refer to python code for the process of stochastic gradient decent.

Using and 80 iterations with standard gradient decent, the cost is 2104853614.71

Using and 3 iterations with stochastic gradient decent, the cost is around the previous number, mostly the cost is smaller than the previous one.

Computation Time: 0.0011930465698242188 (stochastic)

0.017325878143310547 (standard)

Problem 3

1,

2,

1. Proof:

So, the function is convex.

1. Proof:

We need to show is convex, given is convex

Proof done

1. From proof from (a) and (b), we have the new loss function:

is also a convex function.