2. Cost Function

August 5, 2021

1 Hypothesis

The hypothesis function will generally be linear (for simplicity) and thus takes on the following form $h_\theta(x) = \theta_0 + \theta_1 \times x$ where θ is know as a parameter.

The objective in defining a hypothesis function is to find values for θ so that the straight line fits the data well, where the data represents the inputs and outputs the function is suppose to predict.

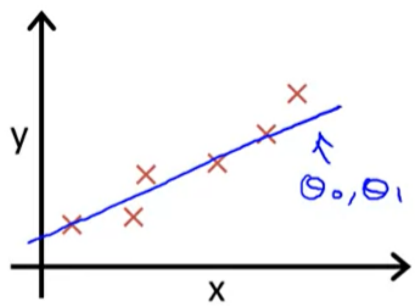


Figure 1: Linear function that approximately fits the input y against the output y

To find values for the parameters that is a good fit to the data, we choose values so that $h_\theta(x)$ is close to y for our training examples (x, y), this boils down to a minimization problem (to minimize the error between the hypothesis function and the training examples)

2 Cost Function

The accuracy of the the hypothesis function can be measured using a cost function

The cost function takes the average difference of all the results of the hypothesis given the input of x's and the expected outputs (y's from the training set). Put in mathematical terms

$$
\begin{aligned}
J(\theta_0, \theta_1) &= \tfrac{1}{2m}\textstyle\sum_{i=1}^{m}(\hat{y}_i - y_i)^2 \\
&= \tfrac{1}{2m}\textstyle\sum_{i=1}^{m}(h_\theta(x) - y_i)^2
\end{aligned}
\tag{1}
$$

The square is taken to obtain positive differences, the division by m is in line with obtaining the average of something and the halving of the mean is for convenience when finding the gradient descent of the cost function where the derivative of $(h_\theta(x) - y_i)^2$ cancels with $1\frac{}{2-gradientdescentdiscussionstilltofollow.}$

3 Cost Function Intuition 1

Notice that from the previous section that our hypothesis function $h_\theta(x)$ is a function of x for fixed values of the parameters θ_0 and θ_1

Whereas our cost function $J(\theta_0, \theta_1)$ is a function of our parameters θ_0 and θ_1

Consider an example. Given our hypothesis function $h_\theta(x) = \theta_0 + \theta_1 x$ has $\theta_0 = 0$ for simplicity, and it only fits 3 values $(1,1), (2,2), (3,3)$ our cost function for different values of θ_1 gives us:

$$
\begin{aligned}
J(\theta_0, \theta_1) &= \tfrac{1}{2\times 3}\textstyle\sum_{i=1}^{3}((x^i) - y^i)^2 \\
&= \tfrac{1}{6}[(1-1)^2 + (2-2)^2 + (3-3)^2] \\
&= \tfrac{1}{6} \times 0 \\
&= 0
\end{aligned}
\tag{2}
$$

for $\theta_1 = 1$

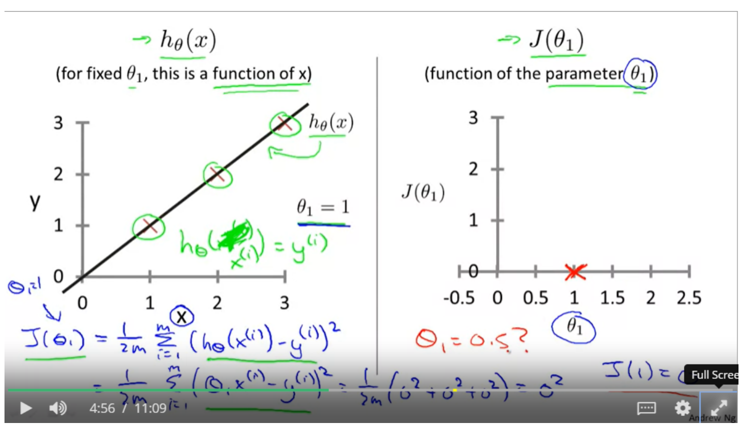


Figure 2: Cost is 0 with $\theta_1 = 1$

setting $\theta_1 = 0.5$

$$
\begin{aligned}
J(\theta_0, \theta_1) &= \tfrac{1}{2\times 3}\textstyle\sum_{i=1}^{3}((x^i) - y^i)^2 \\
&= \tfrac{1}{6}[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2] \\
&= \tfrac{1}{6} \times 3.5 \\
&= 0.53
\end{aligned}
\tag{3}
$$

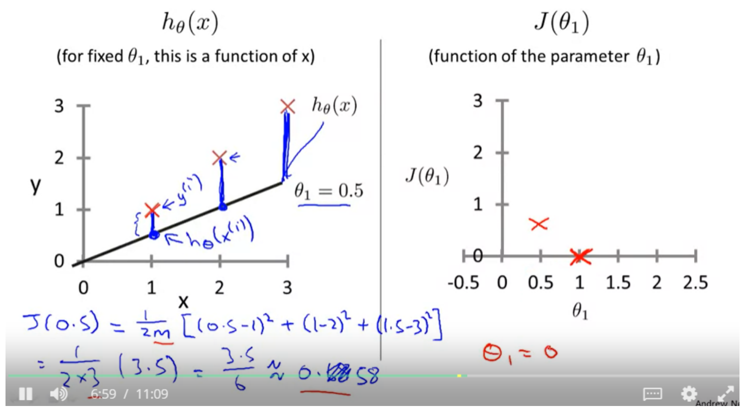


Figure 3: Cost is 0 with $\theta_1 = 0.5$

setting $\theta_1 = 0$

$$
\begin{aligned}
J(\theta_0, \theta_1) &= \tfrac{1}{2\times 3}\textstyle\sum_{i=1}^{3}((x^i) - y^i)^2 \\
&= \tfrac{1}{6}[(0-1)^2 + (0-2)^2 + (0-3)^2] \\
&= \tfrac{1}{6} \times 15 \\
&= 2.5
\end{aligned}
\tag{4}
$$

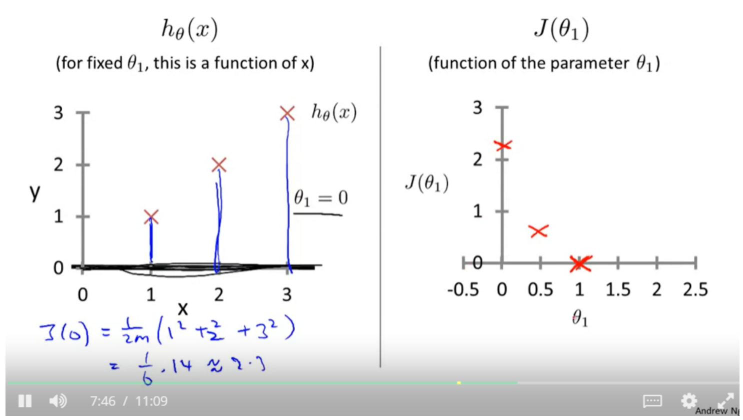


Figure 4: Cost is 0 with $\theta_1 = 0$

carrying on in this manner we can see that $\theta_1 = 1$ is the global minimum of the cost function $J(\theta_1)$

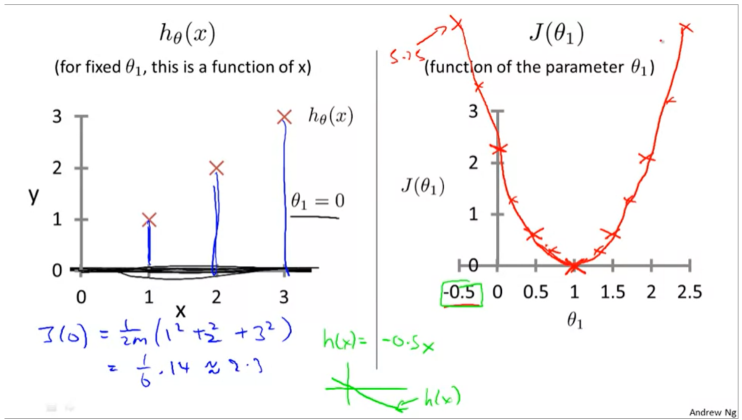


Figure 5: Cost function plot of different values for θ_1

4 Cost Function Intuition 2

The previous intuition section looked at the scenario where θ_0 was set to 0, this made our cost function 2 dimensional. Consider now that we have a non-zero θ_0 value, meaning our cost function is 3D

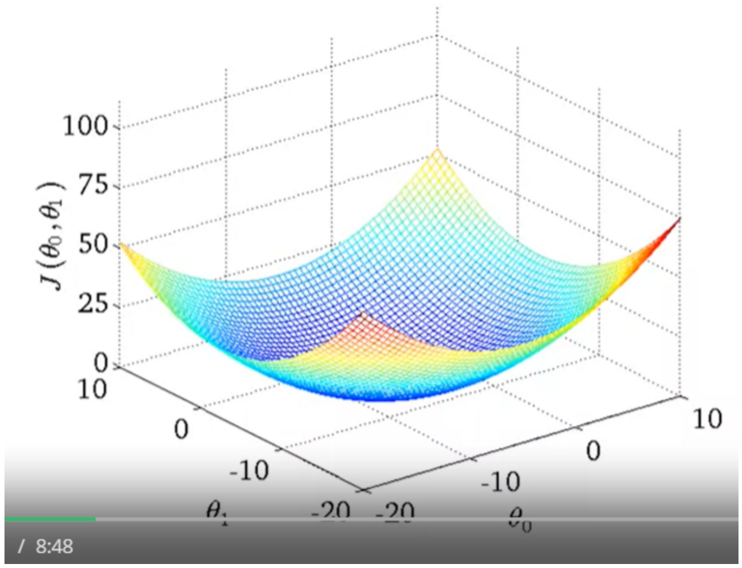
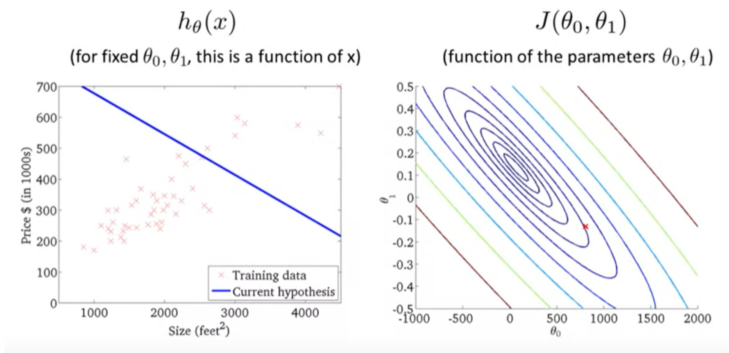
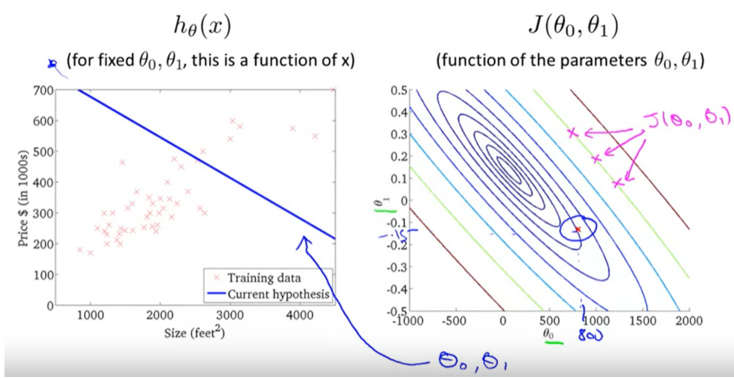
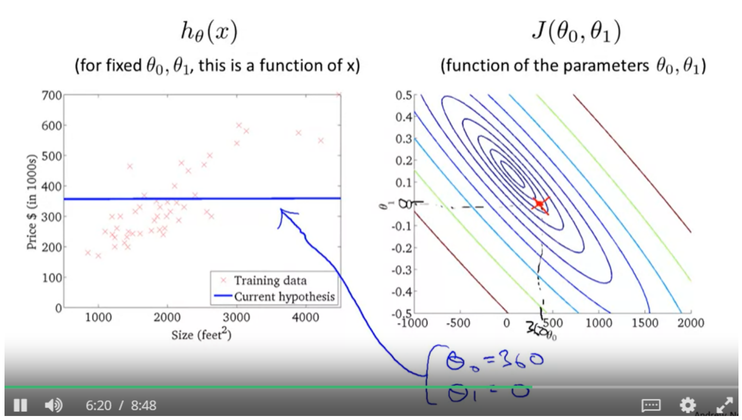


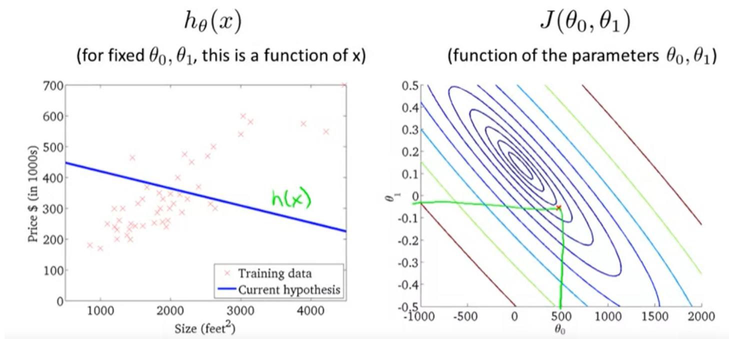
Figure 6: Illustration of the cost function with non zero theta values

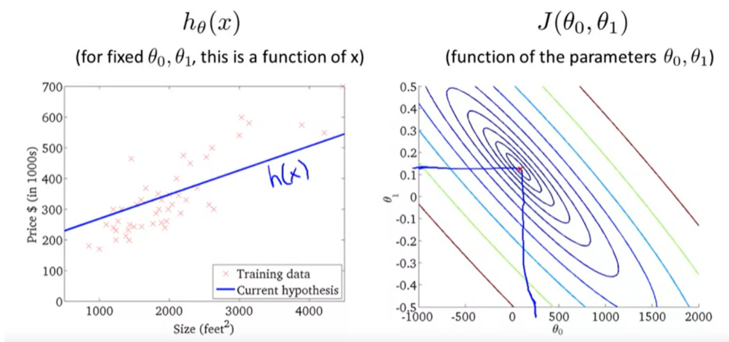
Now, illustrating the effect of different theta values and the effect it has on the hypothesis function but more importantly the cost function we will be making use of contour plots (it is a plane section of the three-dimensional graph - from wikipedia). Each contour corresponds to a different hypothesis function











Notice that we obtain a minimum when the values for θ_0 and θ_1 are approximately $(250, 0.15)$

5 Conclusion

The cost function is a function that tells us how far off our hypothesis function is from predicting the output y of some input x.

The idea with the 2 intuition sections was to show that there exists values for θ_0 and θ_1 defining the hypothesis function such that the cost function is minimized (i.e. the predicted value y given x is close to the real value). Our task is to find these parameter values. We are also not restricted to only 2 parameters but to multiple.