Technical Report 1

1 Linear Regression

We implement linear regression according to the following model

$$h_\theta(x) = \theta^T x = \sum_{i=0}^{n} \theta_i x_i, \tag{1}$$

and the batch gradient descent update rule is

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \tag{2}$$

where $j = 0, 1$ in our case.

After the first iteration, we have

$$\theta_0 = 0.0745, \qquad \theta_1 = 0.3800$$

The gradient descent algorithm runs for about 1500 iteration, and the final values of θ_0 and θ_1 are 0.7502 and 0.0639, respectively. We plot the resulting straight line as well as the given data in Fig. 1

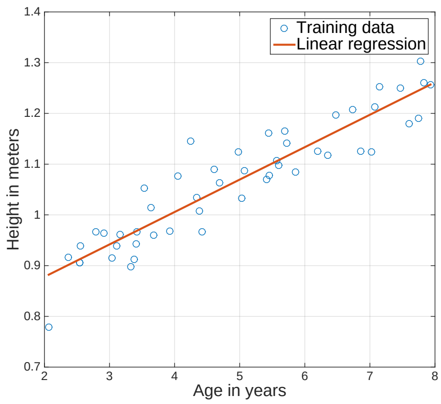

Figure 1: The straight line fitting to the given data.

Being aware of θ, we can make predictions according to Equation (1). For example, given two boys of age 3.5 and age 7, their heights are 0.9737 meters and 1.1975 meters, respectively.

2 Understanding J

The function $J(\theta)$ is visualized by a surface plot, as shown in Fig. 2. It is demonstrated that, $J(\theta)$ has a global minimum with respect to θ.

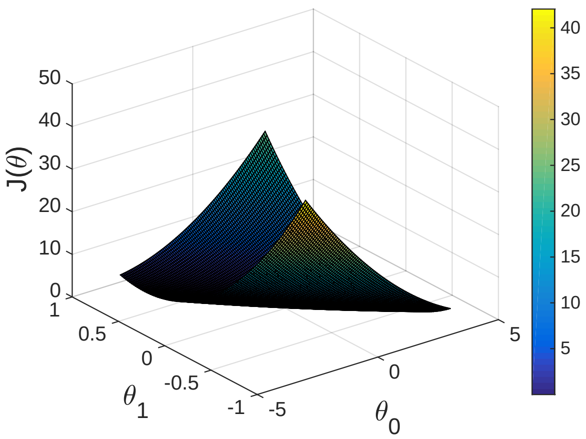

Figure 2: Visualizing $J(\theta)$.

To illustrate the relationship between $J(\theta)$ and θ, we also visualize $J(\theta)$ by a contour plot (see Fig. 3. It is shown that, $J(\theta)$ approaches its minimum when $\theta_0 = 0.7502$ and $\theta_1 = 0.0639$

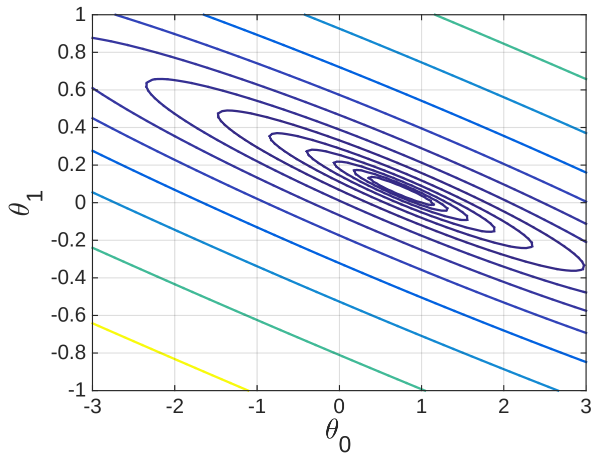

Figure 3: Contour plot of $J(\theta)$.