Foundations of Machine Learning Homework 1

July 27, 2018

2.2.1 Write $J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x_i) - y_i)^2$ as a matrix/vector expression

Let $e = h_\theta(x) - y$ such that $J(\theta) = \frac{1}{m} \cdot e^T e$

2.2.2 Write down an expression for the gradient of J (again, as a matrix/vector expression, without using an explicit summation sign).

Given that $J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x_i)-y_i)^2 = \frac{1}{m}(h_\theta(x)-y)^T(h_\theta(x)-y)$ and that $h_\theta(x) = \theta^T x$, we have that $\nabla J(\theta) = \frac{2}{m} \cdot x^T \cdot (h_\theta(x) - y)$

2.2.3 Use the gradient to write down an approximate expression for the change in objective function value $J(\theta + \eta h) - J(\theta)$

A linear approximation is of the form: $f(x) \approx f(a) + f'(a)(x - a)$

Given the above function value we have the equation of the following form:

$$J(\theta + \eta h) \approx J(\theta) + \nabla J(\theta)(\eta h)$$

Therefore, we have that $J(\theta + \eta h) - J(\theta) \approx \frac{2}{m} \cdot x^T \cdot (h_\theta(x) - y) \cdot (\eta h)$

2.2.4 Write down the expression for updating θ in the gradient descent algorithm. Let η be the step size.

$$\theta = \theta - \eta \nabla J(\theta)$$