ML 2017/2018: Graded Assignment 1 (5 points)

Deadline: October 10 - by Konrad Krawczyk

1. This question is about vectorization, i.e. writing expressions in matrix vector form. The goal is to vectorize the update rule for multivariate linear regression.

(a) Let θ be the parameter vector θ = (θ0 θ1 ···θn) T and let the i-th data vector be: x (i) = (x0 x1 ···xn) T where x0 = 1. What is the vectorial expression for the hypothesis function hθ(x)?

If we have multiple variables, a hypothesis function can be expressed as:

hθ(x) = θ0x0 + θ1x1 + θ2x2 + θ3x3 + θ4x4 (...)

In vectorized form, both the parameter vector θ and the feature vector x are column vectors. If we transpose the parameter vector, it will be possible to multiply θ and x and obtain a single value from this operation. Therefore:

$$h_\theta(x) = \theta^T x$$

(b) What is the vectorized expression for the cost function: J(θ) (still using the explicit summation over all training examples).

This is the default state of the cost function:

$$J(\theta_0, \theta_1, \ldots, \theta_n) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

$$\vec{e} = \sum (h_\theta(x^{(i)}) - y^{(i)})^2 = \sum \ (h_\theta(x)^T - y)^2$$

(we calculate the sum of squared errors)

$$J = \frac{\vec{e}}{2m}$$

(Cost function will be the s.s.e divided by 2*sample size)

(c) What is the vectorized expression for the gradient of the cost function:

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

(d) What is the vectorized expression for the θ update rule in the gradient descent procedure?

$$\theta \coloneqq \theta - \alpha \partial$$

Theta decreases by learning rate multiplied by the gradient(slope) of the cost function.

2. 2 points Derive an equation that can be used to find the optimal value of the parameter θ1 for univariate linear regression without doing gradient descent. This can be done by setting the value of the derivative equal to 0. You may assume that the value of θ0 is fixed.

$$\frac{\partial J(\theta)}{\partial \theta} = J(\theta_0 \theta_1)^T \begin{bmatrix} x_0^{(1)} x_1^{(1)} \\ x_0^{(2)} x_1^{(2)} \\ \cdots \\ x_0^{(m)} x_1^{(m)} \end{bmatrix} = \begin{bmatrix} 0\, 0 \end{bmatrix}$$