Graded Assignment I (part2)

1.

(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)?

h = g(Xθ)

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

J(θ) = (1 / 2m) * (g(Xθ) - $\vec{y}$)

(c) What is the vectorized expression for the gradient of the cost function, i.e. what is:

∂ J(θ) / ∂ θ = (1 / m) * X^T (g(Xθ) - $\vec{y}$)

Again the explicit summation over the data vectors from the learning set is allowed here.

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

θ = θ - (1 / m) * X^T (g(Xθ) - $\vec{y}$)

(e) (bonus points) Vectorization can be taken one step further. We can remove the explicit summation over the training samples by 'hiding´ it in a matrix vector multiplication. Start by collecting all training samples in a data matrix X such that every row of X is a vector from the training set (with the augmented x0 = 1 elements, i.e. the first column of X has elements equal to 1).

??

2. 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.

∂ J(θ_1) / ∂ θ_1 = 0,

θ0 = fixed (constant)

??? //what am I taking the derivative of?