Written part Assignment 1 Machine Learning

Charlotte Out

11426632

Charlotte\_out@hotmail.com

*Exercise 1*

1. Let be the parameter vector and let the i-th data vector be: where = 1. What is the vectorial expression for the hypothesis function

The hypothesis function should look like this

right now, is a (,n) vector, and x is a (,n) matrix too. In order to perform vector multiplication, so to get the to the point where the element of index 1 of , is multiplied with the element of index 1 of x, we should take the transpose of the first vector, in this case , .

Therefore, our vectorized expression will look like this:

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

The cost function is defined in the following way

however, we already defined the vectorized expression of the cost function to be the following:

therefore, the vectorized expression of the cost function is following

we have to index x here since we want to compute the hypothesis of the i-th value of x and than subtract the actual value.

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

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

we see that the “first part” stays the same, we only multiply by a different “index” of x every time. Therefore, we need to write x as an vector of size (n,)

the vectorization will thus look like:

where x is an vector

1. What is the vectorized expression for the update rule in the gradient descent procedure?

where represents the learning rate

we saw that we can write  as

so therefore the vectorized expression of the update rule is:

1. We now have all the vectors

for 1 ranging from 0 to n.is a vector of shape (,n). a 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), would look like this

where , so indeed we see that the first column is which is equal to 1. This is a matrix of shape (n,n)

in our “regular” hypothesis function the x vector is of shape (n,) so we perform matrix multiplication on and. In order to calculate all all the hypothesises for to at the same time we perform this matrix multiplication

we call X for convenience

(2.30, chapter linear algebra, deeplearning book).

This is the norm, which measures the size of the vector, so the size of the difference between and )

= =

*Exercise 2*

The regular opdate rule of gradient descend is as follows:

and the we know that the derivative of the gradient descend is

however, in univariate linear regression we know that we can set the derivative equal to 0.

So we obtain the following equation:

= 0 (we multiply both sides by m and devide by x)

= 0

+ - = 0

+ - = 0

+ - = 0

= -

so indeed, we can calculate without doing gradient descend