## 1 Linear Regression

### 1.1 Derivation of regularized Linear Regression

#### 1.1.1 Preliminary questions

1. The design matrix X is derived from the Error function of polynomial functions for writing it as a product of a vector and a matrix. In this error function we need to find a partial derivate for , for getting equations. The fact that we introduce an x0 (for convention = 1) involves that the matrix (and tor will have a size of m x (n +1) and in the case of the vector of 1 x (n +1).
2. If we consider the cost function , the gradient represents a vector containing the partial derivative of the cost function with respect to every component of the vector and it can be calculated as follows.

Let

Then

The result is an columns vector.

If we want to give a meaning, the gradient denotes the direction of greatest change in the cost function.[[1]](#footnote-1) To minimize the cost function (e.g. find the best parameters ), the gradient must be set equal to zero.

1. The Jacobian matrix is a generalization of the gradient. In general we talk about gradient when we are working on functions that maps , so the Jacobian matrix in this case would just be a vector. We talk about Jacobian matrix when working on functions mapping .[[2]](#footnote-2)

If we have a vector function that maps , where

The Jacobian matrix is defined as

1. The Jacobian matrix is a rows, columns matrix of the form

The first row of the matrix contains the partial derivative of the first element of the hypothesis vector w.r.t. each of the parameters contained in the vector .

#### 1.1.2 Regularized linear regression optimal parameters

The cost function for the regularized linear regression is defined as follows.

We want to show that

Where represents the best fitting parameters to minimize the cost function.

We start by taking the partial derivative of with respect to .

Here we used the fact that .

Now we want to set and isolate .

If we use the fact that , we get

1. <https://en.wikipedia.org/wiki/Gradient> [↑](#footnote-ref-1)
2. <https://math.stackexchange.com/questions/1519367/difference-between-gradient-and-jacobian> [↑](#footnote-ref-2)