Assignment A3

MASD 2018
Department of Computer Science
University of Copenhagen

Casper Lisager Frandsen <fsn483@alumni.ku.dk>

Version 1
Due: December 18th, 23:59

Contents

1. Exercise **3**
(a) . . . 3

2. Exercise **3**
(a) . . . 3
(b) . . . 3

3. Exercise **4**
(a) . . . 4

1. Exercise

(a)

All of the smaller exercises in Exercise 1 have been completed in the attached Jupyter Notebook named "madA3.ipynb"

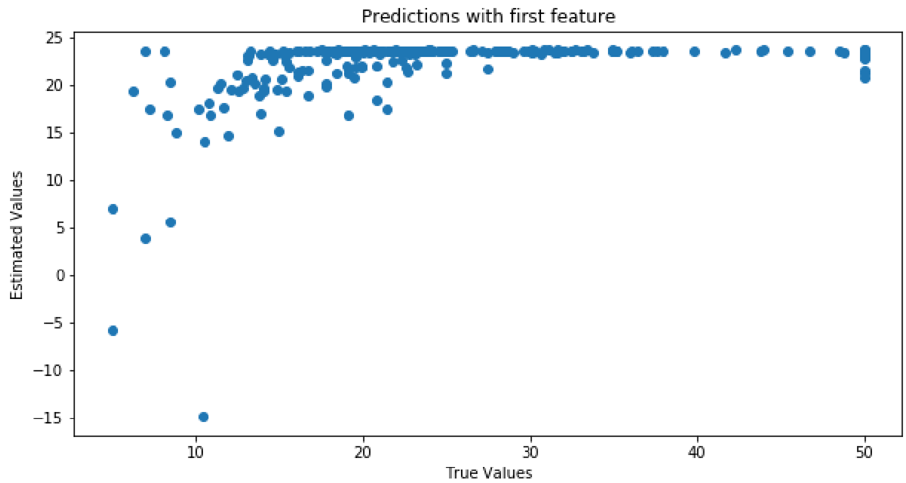
2. Exercise

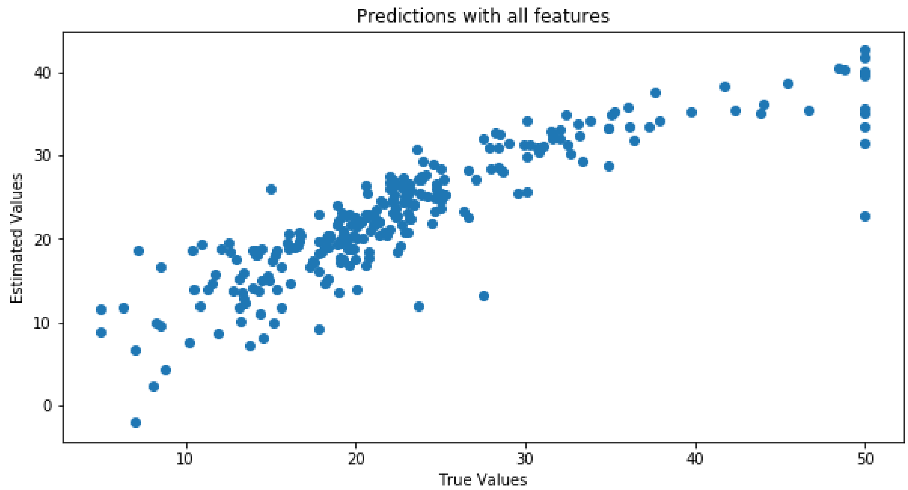
(a)

Most of this exercise has been completed in the attached Jupyter Notebook named "madA3.ipynb", not the "housing_1.py" or "housing_2.py" files. The "linreg.py" file has been modified as per the assignment. Parts of the code have been copied from the lecture slides on Absalon.

(b)

The weights have been written in the attached Jupyter Notebook. We can see that the weights each mean something different: The first weight is the lowest value of the house, and the second weight shows the effect of the crime rate on housing prices.





3. Exercise

(a)

To do the derivation, we will first simplify the expression, by rewriting it to an expression using only matrices and vectors:

$$\nabla \mathcal{L}(w) = \nabla \sum_{n=1}^{N} (f(x_n; w) - t_n)^2 = \nabla \sum_{n=1}^{N} (w^T x_n - t_n)^2$$

$$\nabla (Xw - t)^T (Xw - t)$$

We simplify this further, which enables us to use simpler rules of derivation:

$$\nabla ((Xw)^T - t^T)(Xw - t)$$

$$\nabla (Xw)^T Xw - (Xw)^T t - t^T Xw + t^T t$$

Since $(Xw)^T t$ and $t^T Xw$ both result in single numbers, we can say that they are equivalent since they eventually result in the same calculations:

$$\nabla w^T X^T Xw - 2(w^t X^T t) + t^T t$$

This cannot be simplified further, and is now simple to derive, using the following: $\nabla w^T Cw = 2Cw$, $\nabla w^T x = x$

$$2(X^T Xw) - 2(X^T t)$$

Now we have to set this $= 0$, so we can find the minimum:

$$2(X^T Xw) - 2(X^T t) = 0$$

$$X^T X w = X^T t$$

We now have to multiply both sides with $(X^T X)^{-1}$:

$$w = (X^T X)^{-1} X^T t$$

We can now see that this is the optimal least squares parameter for $\hat{w}$