MAD Assignment 1

Ask Jensen

29. november 2021

Indhold

1 Problem 1

the exercise is to compute the partial derivations of the following three problems each containing a different function.

1.1 (a)

Considering the function, it is fairly straight forward to compute the derivative with respect to x and y. Both derivations will be done using the power rule: $x^n = nx^{n-1}$

$$f(x,y) = x^4y^3 + x^5 - e^y$$

Answer:
calculating $\frac{\partial}{\partial x}$

$$f'_x = 4x^3y^3 + 5x^4$$

calculating $\frac{\partial}{\partial y}$

$$f'_y = x^4 3y^2 - e^y$$

It is worth mentioning, that the exponent (e^y) will derive to itself even though the power function states otherwise.

1.2 (b)

This function is bit harder to compute. I will be using the power rule: $x^n = nx^{n-1}$ and the chain rule: $f(g(x)) = f'(g(x)) \cdot g'(x)$

$$f(x,y) = \frac{1}{\sqrt{x^3 + xy + y^2}}$$

I will start by deriving the innter function with respect to x and y since these derivations are going to be used later on in the calculations.

calculating $\frac{\partial}{\partial x}$

$$f'_x = 3x^2 + y$$

calculating $\frac{\partial}{\partial y}$

$$f'_y = x + 2y$$

Answer:
I will rewrite the function using simple fraction rules in order to make the function a bit easier to work with.

$$f(x,y) = (x^3 + xy + y^2)^{-\frac{1}{2}}$$

applying the power rule

$$f(x,y) = -\frac{1}{2}(x^3 + xy + y^2)^{-\frac{3}{2}}$$

applying the chain rule

$$f(x,y) = -\frac{1}{2}(x^3+xy+y^2)^{-\frac{3}{2}} \cdot (3x^2+y)$$

calculating $\frac{\partial}{\partial x}$

$$f'_x = -\frac{3x^2+y}{2(x^3+xy+y^2)^{\frac{3}{2}}}$$

I will not be showing the full calculation for f'_y since it derives trivially in the same was as f'_x, with the only exception being that instead of using the chain rule with the derivation of x you now use the derivative of y thus I will only show the last two steps

applying the chain rule

$$f(x,y) = -\frac{1}{2}(x^3+xy+y^2)^{-\frac{3}{2}} \cdot (x^4 3y^2 - e^y)$$

calculating $\frac{\partial}{\partial x}$

$$f'_x = -\frac{x+2y}{2(x^3+xy+y^2)^{\frac{3}{2}}}$$

1.3 (c)

I order to compute this function I will be using the power rule: $x^n = nx^{n-1}$ and the quotient rule: $(\frac{f(x)}{g(x)})' = \frac{g(x)f'(x)-f(x)g'(x)}{g(x)^2}$

$$f(x,y) = \frac{x^3+y^2}{x+y}$$

Answer:
I will start by finding the derivations for f(x) and f(y) aswell for g(x) and g(y). these calulations are done using the power rule. Having these calculating done, it's a simple matter of substituting with the "new"expressions in the qoutient rule
$f'_x \cdot g(x) = 3x^2(x+y)$
$f'_y \cdot g(y) = 2y(x+y)$
$f(x) \cdot g'_x = 1 \cdot (x^3+y^2)$
$f(y) \cdot g'_y = 1 \cdot (x^3+y^2)$

$$f'_x = \frac{3x^2(x+y)-(x^3+y^2)}{(x+y)^2}$$

$$f'_y = \frac{2y(x+y))-(x^3+y^2)}{(x+y)^2}$$

2 Problem 2

In order to solve these exercises I will make use of *Tabel 1.4* in the book "A first couse in machine learning"

2.1 (a)

$f(\overline{x}) = \overline{x}^T\overline{x} + c$
Answer:
Here we have the "third case"of the table, $\overline{x}^T\overline{x} \Rightarrow \nabla f(\overline{x}) = 2\overline{x}$
The constant c is simply discarded, when differentiating.

2.2 (b)

$f(\overline{x}) = \overline{x}^T\overline{b}$
Answer:
Here we have the "first case"of the table, $\overline{x}^T\overline{b} \Rightarrow \nabla f(\overline{x}) = \overline{b}$

2.3 (c)

$f(\overline{x}) = \overline{x}^T A\overline{x} + \overline{b}^T\overline{x} + c$
Answer:
Here we have the a combination of "fourth case"and the "second case"from the table.
$\overline{x}^T A\overline{x} \Rightarrow 2A\overline{x}$
and
$\overline{b}^T\overline{x} \Rightarrow \overline{b}$
Combining the two, we get the result: $\nabla f(\overline{x}) = 2A\overline{x} + \overline{b}$
The constant c is simply discarded, when differentiating.

3 Problem 3

3.1 (a)

Computing the "naive"mean using the training set.

$$mean = 22.016601$$

this value is used to obtain the estimates for a new house

3.2 (b)

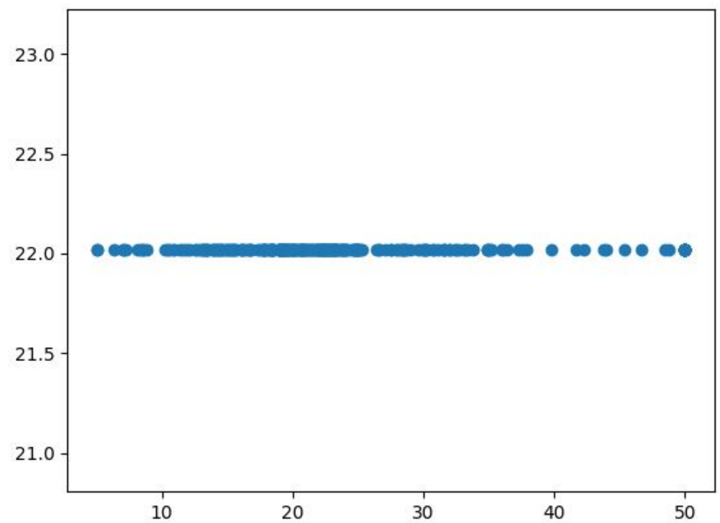
calculating the Root Mean Sqare Error value.

$$RMSE = 9.672478$$

3.3 (c)

For the rest of the code/results see the extended **housing_1.py**

The figure shows the 2d scatter plot for ("True house prices"vs. "Estimates") for all instances in the test set



4 Problem 4

4.1 (b)

Fitting on the first feature (CRIM) the fit() function gives me with the following weights

$$w_0 = 23.63506195$$

and

$$w_1 = -0.43279318$$

These values tell something about the context about the houseprices and the crime-rate. The value for a house is approximately \$23.000 and as the crime rate rises, the value of the houses will go down by a factore of -0.43279318

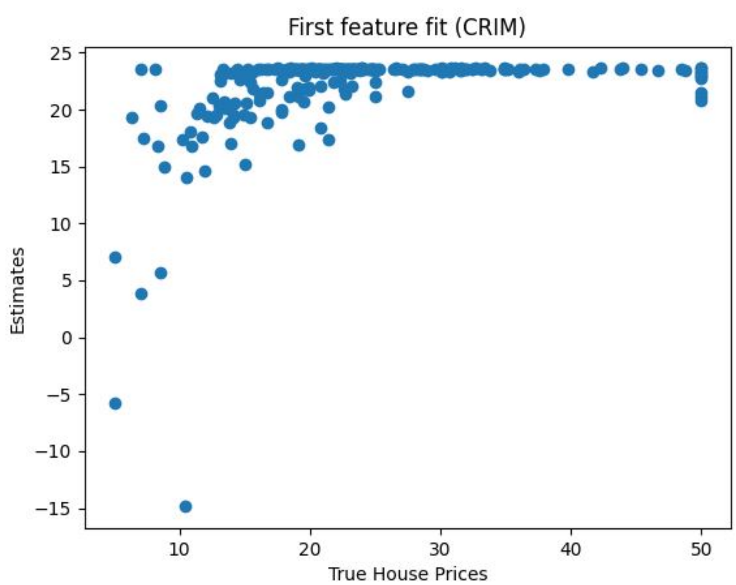
4.2 (c)

Fitting on alle the features the fit() function gives me with the following weights following weights.

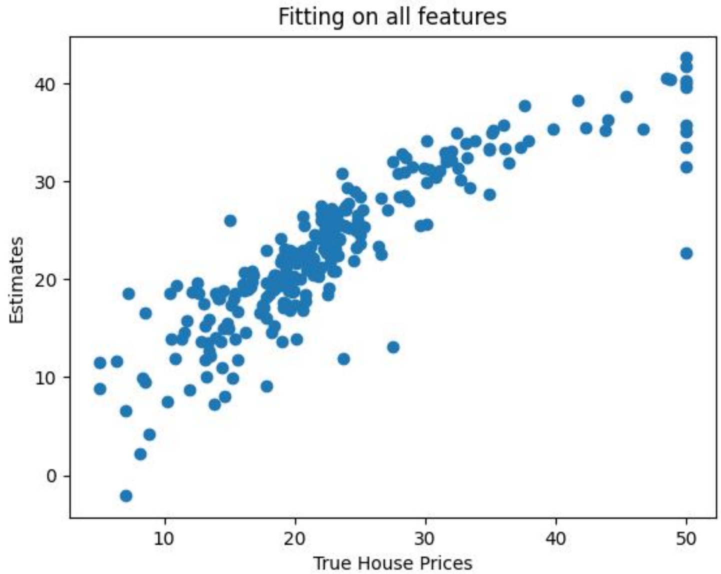
$$\hat{\mathbf{w}} = \begin{bmatrix} 2.05558292e+01 \\ -2.50312466e-02 \\ 1.73555758e-02 \\ 1.80262164e-01 \\ -1.40859425e+00 \\ 7.55741436e+00 \\ -1.86729639e+00 \\ 1.90656514e-02 \\ 9.81506395e-01 \\ -9.99365107e-02 \\ 4.17706547e-03 \\ 2.56978545e-01 \\ -1.43515875e-03 \\ -2.36790752e-01 \end{bmatrix} \tag{1}$$

4.3 (d)

Calculating the RMSE
$RMSE_{singlefeature} = 8.954859906611233$
$RMSE_{allfeatures} = 4.688333653627553$
What we can learn from the two RMSE values is, that it is possible to obtain a better fit when using more features.



Figur 1: House prices for one feature



Figur 2: House prices for all features

For the rest of the code/results see the extended **housing_2.py** file that has been uploaded.

5 Problem 5

$$\begin{aligned}\mathcal{L} &= \sum_{n=1}^{N} (\mathbf{w}^T \mathbf{x}_n - t_n)^2 \\ &= (\mathbf{X}\mathbf{w} - t)^T (\mathbf{X}\mathbf{w} - t)^T \\ &= ((\mathbf{X}\mathbf{w})^T - t^T)(\mathbf{X}\mathbf{w} - t)^T \\ &= \mathbf{w}^T \mathbf{X}^T \mathbf{X}\mathbf{w} - 2\mathbf{w}^T \mathbf{X}^T t + t^T t\end{aligned}$$

using "case 4"from the table 1.14

$$\mathbf{w}^T \mathbf{X}^T \mathbf{X}\mathbf{w} = 2\mathbf{X}^T \mathbf{X}\mathbf{w}$$

and "case 1"gives me

$$2\mathbf{w}^T \mathbf{X}^T t = 2\mathbf{X}^T t$$

thus

$$\frac{\partial \mathcal{L}}{\partial w} = 2\mathbf{X}^T \mathbf{X}\mathbf{w} - 2\mathbf{X}^T t = 0$$

+2 and -2 cancel out leaving me with

$$\mathbf{X}^T\mathbf{X}\mathbf{w} - \mathbf{X}^T t = 0$$

$$\mathbf{X}^T\mathbf{X}\mathbf{w} = \mathbf{X}^T t$$

multiplying both sides with $(\mathbf{X}^T\mathbf{X})^{-1}$ which can be rewritten as $\mathbf{I}$ which is the denotation of the Identity matrix

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

Multiplying the vector $\mathbf{w}$ with the Identity matrix, will simply return the vector $\mathbf{w}$ thus the result is

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

The total training loss is the average loss multiplyed by the number of instances in the dataset. which is excatily the same as the value we obtain from the avereage loss function