ML-HW1

Problem 4: Programming: prediction of acute aquatic toxicity

(a) implement the stochastic and batch gradient descent algorithms as well as the analytical solution. Compare the obtained estimates with those of the built-in function lm in R.

lm : a function for Fitting Linear Models in R

```
lm(formula = data$y ~ data$x, data = data)

##Coefficients:
(Intercept)        data$x
      2.6683        0.6619
```

ALF: My analytical function for Fitting Linear Models in R

```
ALF(mean_data,3)

##"y= 2.66828890038184 + 0.661902756614365 * x"
    4.653997
```

BGD_LF: batch gradient descent function for Fitting Linear Models in R

```
> BGD_LF(data,3,0.0001)

##"y= 2.66828890038044 + 0.661902756614828 * x"
   4.653997
```

SGD_LF: stochastic gradient descent function for Fitting Linear Models in R

```
> SGD_LF(data,3,0.0001)

[1] "y= 2.55570130195476 + 0.707942624628499 * x"
[1]    4.679529
```

(b) implement KNN and Nadaraya-Watson kernel regression without using existing packages and apply it to the above dataset.

```
> Nadaraya_watson(data,3,4)

[1] 4.227973

> KNN(data,3,3)

[1] 4.313
```

(c) randomly split the data into training and test sets and compare the performance of the above models.

```
> test_error_ALF
[1] 123.887

> test_error_SGD
[1] 325.372

> test_error_BGD
[1] 126.0875

> test_error_KNN
[1] 137.9366

> test_error_nadaraya
[1] 155.5612
```

Between all of above models, it seems analytical solution has the least test error.