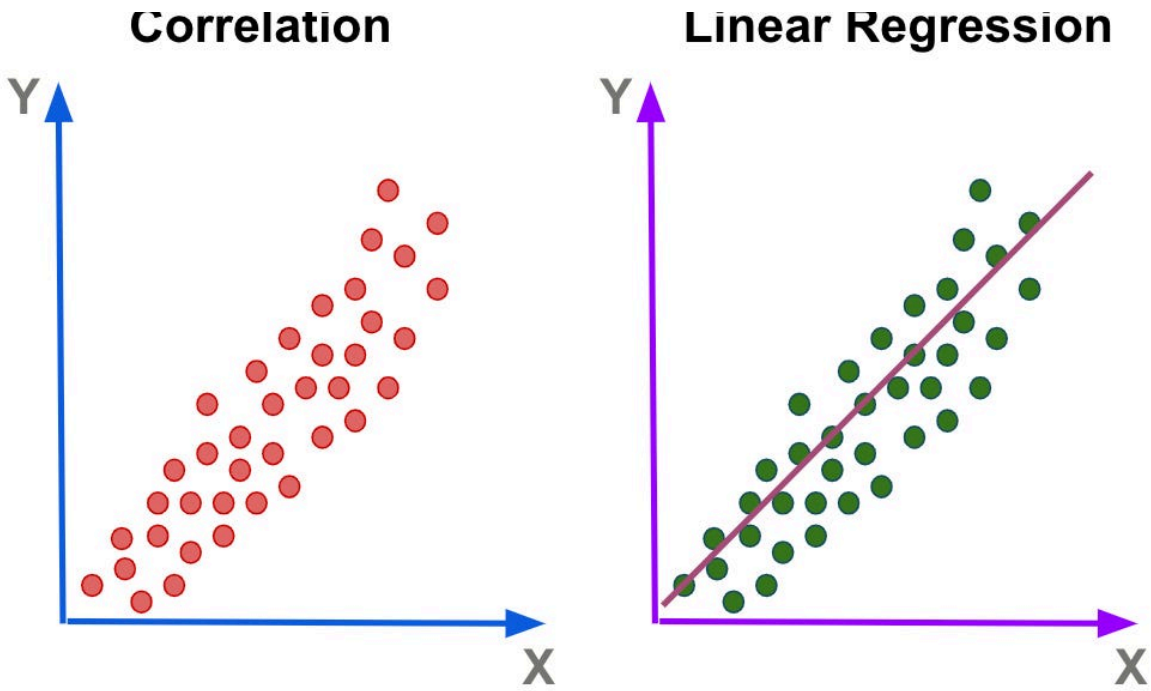
Generative Models

Consider the problem of linear regression where the data points have one independent variable and another dependent variable. Your task is to find a linear function of dependent variable which best fit the observed values of the dependent variable.



Here Y - dependent on x.
As we can see that the linear function intuitively represents the trend in the values of Y for different values of X.

Check Point 1. Understand the linear regression problem statement carefully from https://www.geeksforgeeks.org/ml-linear-regression/ or https://www.cse.iitb.ac.in/~swaprava/cs217240_2024.html

We all know that this data originated from the environment. But it turns out that there is an amazing correlation between Y and X. With increasing X, Y observed is also increasing.

To think about such situations we use generative models. A generative model is an experiment model which tries to explain how this data is generated hypothetically. Let's understand this in the above context.

If I say that this data is not generated from the environment rather it has been obtained by a boy who first chooses a real number(say x), then he multiplies this number by **w** and add **b**. So the value obtained is **w.**x + **b.** Then this boy generates a random number from a gaussian distribution with mean 0 and variance **v**. And finally he adds this random value to our above computed linear function of x and then he presents in front of you the x as dependent variable and the sum of random + computed as independent variable.

Take a pause and think about this.

What we have presented right now is a hypothetical experiment from which we **expect/believe** that the data is generated by running it multiple times. What we don't know is that what are the parameters.

Check Point 2. Implement an experiment in python which first uniformly generates a number from 0 to 10 (need not be integer) and does the above thing with **w = 1, b = 0 and v = 1.** Obtain 1000 points in this way and plot the observed points using matplotlib.pyplot.

Questions : What kind of pattern did you obtain in the observed points? Can you relate this with the above data? What can be the uses of generative models according to you ?

Objective Functions

For now assume that it is indeed the case that the data is generated from our believed generative model. How to find the values of parameters like in above case w,b and v?

For that what we use is a cost function which gives a number to the goodness of our current parameters. Then we try to find the parameter values which minimizes the cost function.

Check Point 3. Go through this to understand parameter estimation.
https://web.stanford.edu/class/archive/cs/cs109/cs109.1192/reader/11%20Parameter%20Estimation.pdf

Questions : Write a short report based on your understanding of MLE.

Understanding the generative model of our problem.

https://drive.google.com/file/d/1oDeqXBFYn2WZeY-27UqZmDwsvpnX77fl/view

Check Point 4. Go through the 36 - 39 numbered pages to get what our generative model is. Feel free to contact me if you have any difficulties in understanding this generative model.

Questions: Write down a short report explaining our generative model.

Understanding our objective function

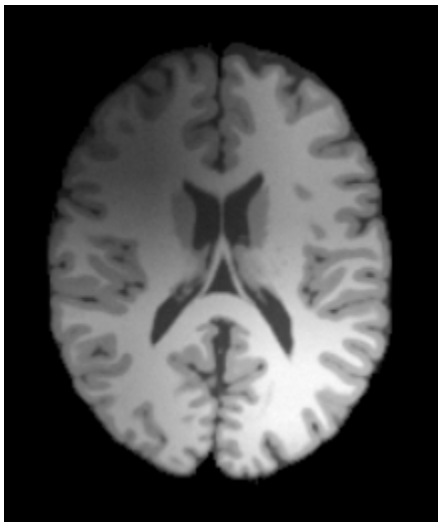
For understanding this you need have idea of fuzzy membership.
Check Point 5. Go through the 23 - 35 numbered pages to understand Fuzzy C Means. (No need to implement this 🥳).

Now you have to understand our objective function.
Check Point 6. Go through the 39 - 45 numbered pages.

Questions: I will have an individual meet with all of you to check if you have understood it or not.

Implementing our algorithm

Consider the image below.



You have to implement the algorithm in python using only basic libraries - **Numpy**, **Pandas**, **Matplotlib**. You have to take into consideration the **masking** of the **background.**

Your code should be runnable and well commented.
You have to finally return the heat map of the bias field.
I would like you to tune the neighborhood, weights etc. in order to get a good result.

You should think about what kind of impact these things have. This part will also test how deeply you understand the algorithm and its various aspects. Create a log of whatever you try and add it to the report.

A final note

All the best 😁