Modeling a Problem with Linear Regression.

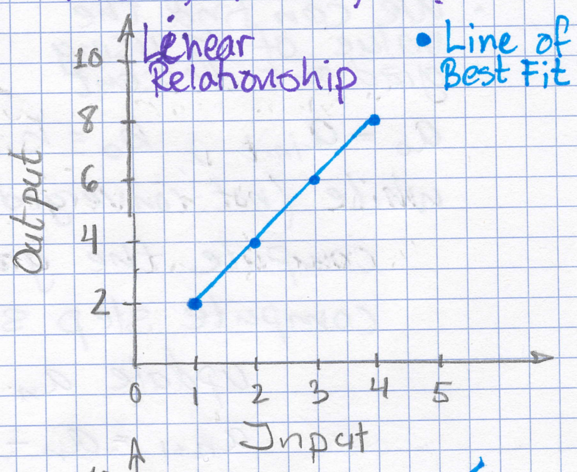
- A brief overview of Linear Regression
- How can we implement Linear Regression
- Experimenting with the learning factor
- What if there is no linear relationship

A brief overview of Linear Regression

- Creating a line of best fit to the data
- The aim is to predict an output of a linear relationship for a certain input

How Do we Create a Line of Best fit?

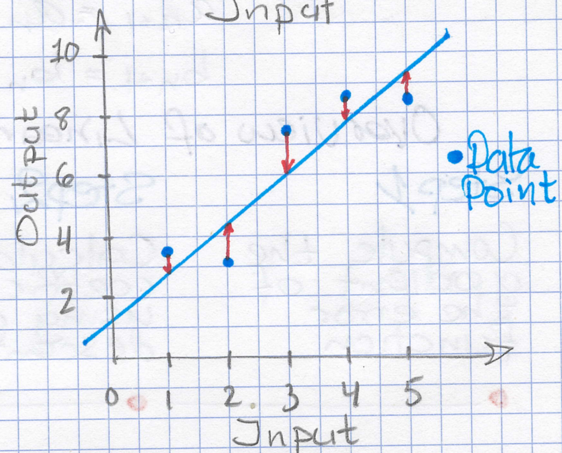
- The line of Best fit is of the mathematical form - $f(x) = ax + b$
- We obtain the line of best fit through optimisation where we are minimising the error between the data points and the line of best fit



Least Squares

- Least squares minimizes the square of distance from the line to a data point.
- The error function we are aiming to minimise is defined as:

$$\text{Error}(a,b) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - (ax_i + b)\right)^2$$

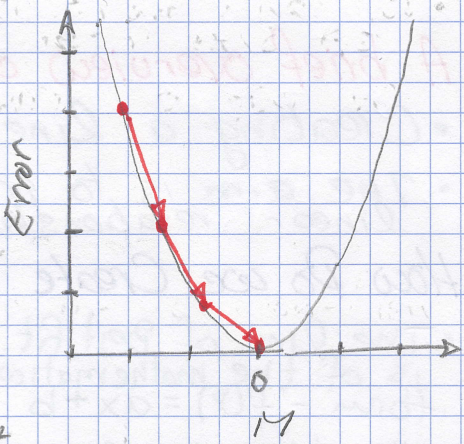


- The minimum is obtained when the gradient is equal to zero.

$$\nabla_a E = \frac{2}{N}\sum_{i=1}^{N} -x_i(y_i - (ax_i + b))$$

$$\nabla_b E = \frac{2}{N}\sum_{i=1}^{N} -(y_i - (ax_i + b))$$

Gradient Descent



- We want to find the value of M that minimises our error, this is achieved when the gradient is zero.
- We can find the value of M using gradient descent

$a_0 = a_{init}$, $b_0 = b_{init}$

while (not converged) do following:

compute the gradient for g_a and g_b

compute step size. $\alpha = 1/n + 2$

update a_n and b_n

$a_{n+1} = a_n - \alpha g_a$

$b_{n+1} = b_n - \alpha g_b$

Overview of Linear Regression.

Step 1

Compute the gradient of the error function

Step 2

Calculate the coefficients using gradient descent

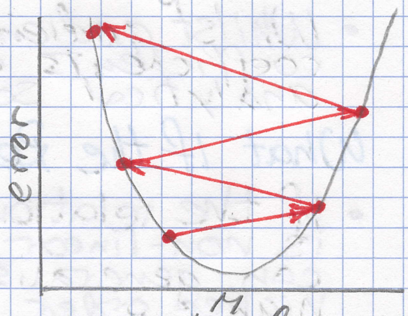
Step 3

Run the algorithm to test the results

How do we calculate the step size?

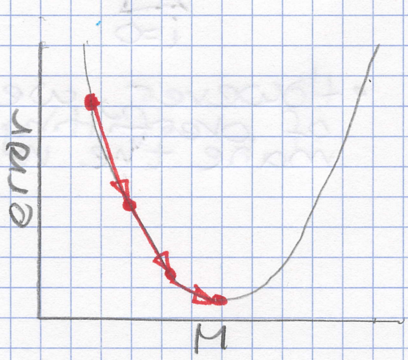
- We can keep the step size constant. However if it's too lage the solution can diverge
- We can use backtracking line search. This is where we decrease the step size by a certain amount at each iteration.
- We can compute an approximate of the step size from a formula.
- Or we can compute the exact optimum step size

Constant step size



- A constant step size is simple, but it is often not the optimum value
- In some cases, it can lead to the divergence of the solution.
- Ideally we want the step size to decrease

Monotonically decreasing the step size



- Ideally we want the step size to decrease monotonically as the algorithm converges.
- This can be achieved from the following two equations:

$$\alpha = 2/(n+2)$$

$$\alpha_{n+1} = \beta\alpha_n, \beta \in (0,1)$$

- We can also compute the exact step size defined as:

$$\alpha = \mathrm{argmin}_\alpha f(a_{n+1}, b_{n+1})$$

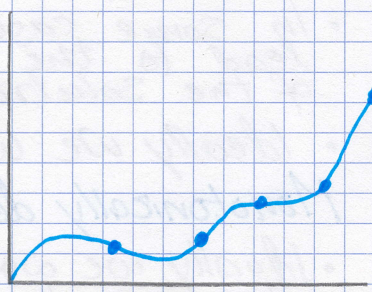
Overview of Linear Regression

- Linear Regression allows us to model a linear relationship which we may not know in advance
- We can generalize linear regression to polynomial regression by including higher order terms.
- Through gradient descent, we obtain the coefficient for a line of best fit.
- The step size determines how fast we converge to a solution.
- Whilst gradient descent is useful, the coefficients can be obtained from an analytical solution.

What If the Relationship is Non-Linear?

- If the relationship is non-linear, we can generalise with higher order polynomial

$$f(x) = \sum_{i=0}^{N} a_i x^i$$



- However, we need to be careful of overfitting, this is where we make the model too complex.