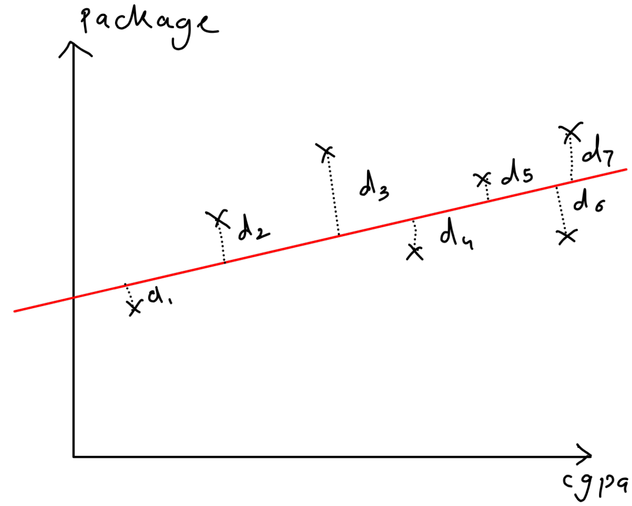
Simple Linear Regression

Two methods to Calculate m and b

OLS (generally used in libraries like Scikit learn)

Gradient descent (Used when there are alot of dimensions)

Ordinary Least Square



Our goal is to find some value of m and b such that the loss function is least.

$\hat{y}$ is called as model's prediction in machine learning

$\hat{y}_i = m x_i + b$ (Prediction)

$d_i = (\underbrace{y_i}_{\text{Real value in the data}} - \underbrace{\hat{y}_i}_{\text{value predicted by the model}})$

Error

$E = d_1 + d_2 + d_3 \dots + d_n$

$= d_1^2 + d_2^2 + \dots + d_n^2$

$\boxed{E = \sum_{i=1}^{n} d_i^2}$ – Error function or Loss function

$$\therefore \boxed{E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

$$E(m,b) = \sum_{i=0}^{n} (y_i - mx_i - b)^2$$

Lets start the math

$E(m) = \sum_{i=1}^{n} (y_i - mx_i)^2$ (Here the y axis is constant and the slope m is not)

$E(b) = \sum_{i=1}^{n} (y_i - x_i - b)^2$ (Here the slope m (angle) is constant. It will only move up and down (y axis))

Here our main goal is to find the perfect value for m and b such that the error function is the least. To do this, we use the concept of maxima and minima. Therefore we will derivate the function.

$$\frac{\partial E}{\partial b} = \frac{\partial}{\partial b}\sum_{i=1}^{n}(y_i - mx_i - b)^2 = 0$$

$$= \sum \frac{\partial}{\partial b}(y_i - mx_i - b)^2 = 0$$

$$= \sum (-2)(y_i - mx_i - b) = 0$$

$$= \sum y_i - mx_i - b = 0$$

Mean

$$\frac{\sum y_i}{n} - \frac{\sum m x_i}{n} - \frac{\sum b}{n} = \frac{0}{n}$$

$$\bar{y} - m\bar{x} - \cancel{n}\frac{b}{\cancel{n}} = 0$$

$$\bar{y} - m\bar{x} = b$$

$$\therefore \boxed{b = \bar{y} - m\bar{x}}$$

Now, we will equate Error function with using value of b

$$E = \sum(y_i - mx_i - \bar{y} + m\bar{x})^2$$

$$\frac{\partial E}{\partial m} = \sum \frac{\partial}{\partial m}(y_i - mx_i - \bar{y} + m\bar{x})^2 = 0$$

$$= \sum 2(y_i - mx_i - \bar{y} + m\bar{x})(-x_i + \bar{x}) = 0$$

$$= \sum -2(y_i - mx_i - \bar{y} + m\bar{x})(x_i - \bar{x}) = 0$$

$$= \sum (y_i - mx_i - \bar{y} + m\bar{x})(x_i - \bar{x}) = 0$$

$$= \sum [(y_i - \bar{y}) - m(x_i - \bar{x})](x_i - \bar{x}) = 0$$

$$= \sum [(y_i - \bar{y})(x_i - \bar{x}) - m(x_i - \bar{x})^2] = 0$$

$$= \sum (y_i - \bar{y})(x_i - \bar{x}) = m\sum(x_i - \bar{x})^2$$

$$m = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Regression Metrics

① MAE : Mean Absolute Error

It calculates the average of the absolute differences between predictions (y_i) and actual values ($\hat{y}_i$).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} | y_i - \hat{y}_i |$$

n : number of data points
y_i : Actual Value
$\hat{y}_i$: Predicted Value

Advantage

i. Same unit for the output

ii. Robust to outliers.

Disadvantage

i. Because of the modulus, the graph of modulus passes through 0, and differentiating this is a difficult task.

② MSE : Mean Squared Error

MSE calculates the mean of the squared differences between predictions ($\hat{y}_i$) and actual values (y_i).

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

n: Number of data points
y_i : Actual Value
$\hat{y}_i$: Predicted Value

Advantage:

i. Since this is differentiable, this can be used as Loss function.

Disadvantage:

i. Since we are squaring, the output will be squared. So we need to convert it to root.

ii. Not Robust to Outliers.

③ RMSE : Root of Mean Squared error.

$$RMSE = \sqrt{mse}$$

$$= \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

Advantage

It's output will be in the same unit since it's square root of MSE.

④ R2 Score (Coefficient of Determination)

or

: $\boxed{\text{Goodness of fit}}$

Let's take an example.

You have GPA and salary data.

Now if some one ask's you what is the package that he will be having in future (he does not have GPA), we generally find the mean of the salary.

And if someone has GPA, we do regression.

So, R2 score simply calculates the difference in the mean and the actual regression line.

$$R2 = 1 - \frac{SS_r}{SS_m}$$

SS_r : Sum of squared error for regression line

SS_m : Sum of mean line

$$= 1 - \frac{\left[\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right]_{Reg}}{\left[\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right]_m}$$

R2 Score ranges from 0 to 1. The closer to 1, the better the model.

If the score is -ve, it means the regression line (model) has the worst performance.

The more rows added to the data, the less reliable R2 score becomes.
(predictors)

It only determines correlation and not causation.

So, we modify it to improve this flaws

⑤ Adjusted R2 Score.

This is modified version of R2 score that accounts for the number of predictors in a regression model. It adjusts for inculsion of irrelevant predictors, providing a more accurate measure of model fit.

$$R^2_{adj} = 1 - \left(\frac{1 - R^2)(n-1)}{n-p-1} \right)$$

R^2 : R2 Score
n : Number of observations
p : Number of predictors (row)

The main purpose is that it penalizes the addition of irrelevant predictors to prevent overfitting.

∴ This focuses only on meaningful variable that contribute to model.