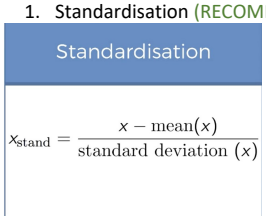
Feature Scaling

Wednesday, March 30, 2022 2:53 PM

Note:
- Feature Scale must be applied after the split and we won't apply feature scale on whole matrix of features 'X' but ofcourse on both X_train and X_test separately. And SCALER will be fitted to only X_train and then we will transform X_test(we will apply feature scaling on X_test because X_test is something that we are not supposed to have during the training but only after the production.
- We are not allowed to fit our feature scaling tool on the test set.

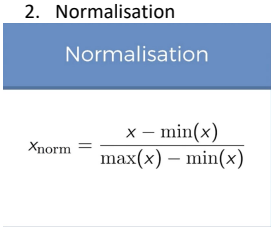
What?:
=>Feature scaling allows to put all our features on the same scale
Why?
=>Inorder to avoid some features to be dominated by other features in such a way that dominated features are not even considered by some Machine Learning models.
Where?
=>We won't have to apply feature scaling for all the Machine Learning models but only for some of them.
How?
=>By using main two feature scaling technique. They are:

1. Standardisation (RECOMMENDED MOST OF THE TIME)



- Standardisation consists of subtracting each value of your feature by mean of all the values of the feature and then dividing by the standard deviation(which is the square root of the variance)
- This will put all the values of the feature between around -3 and +3.

2. Normalisation



- Normalization consists of subtracting each value of your feature by the minimium value of the feature and then dividing by the difference between the maximum value of the feature and the minimum value of the feature.
- All the values of the features will be positive and it will be between 0 and 1.

Should we go for Standardisation or Normalisation?

=>Normalisation is recommended when you have normal distribution in most of your features. This is will be a great feature scale technique in that case.
=>Standardisation actually works well all the time. It does the job all the time. Therefore, this is the technique that will work all the time and this is the technique that is more recommeneded for some specific situation where you have most of the feature following normal distribution and hence, going for standardisation is a great choice. This is always work and this will always improve the training process.

Most Frequently Asked Question in Data Science Community

Do we have to apply feature scaling (Standardisation) to the dummy variables in matrix of features?

=>NO!!!

- Because the goal of Feature Scaling (Standardisation) is to have all the values of the features in the same range and since, standardisation actually transforms your features that they take values between -3 and +3.
- Moreover, Dummy Variable already take values between -3 and +3 (dummy variable is equaled either 1 or 0 so, there is nothing extra to be done here with Standardisation and Standardisation will only make it worse even though it will transform the value between -3 and +3 but we will totally loose the interpretation of these variables and also we will get nonsense numerical values.
- In the other words, we will loose the information of which 'Country'(Refer Data.csv) corresponds to the observation.

Feature Scaling

Standardisation	Normalisation
$x_{stand} = \frac{x - \text{mean}(x)}{\text{standard deviation }(x)}$	$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)}$

Regressions

Simple Linear Regression

$$y = b_0 + b_1 x_1$$

Multiple Linear Regression

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$

Polynomial Linear Regression

$$y = b_0 + b_1 x_1 + b_2 x_1^2 + \ldots + b_n x_1^n$$