Function Transformer

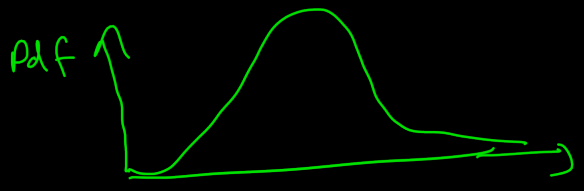
Mathematical Transformations

Feature Engineering
↳ Feature Transformation
↳ Mathematical transformation

1) Log transformation
2) Reciprocal transformation
3) Power (Sqr / Sqrt)

1) Box-cox
2) Yeo-Johnson

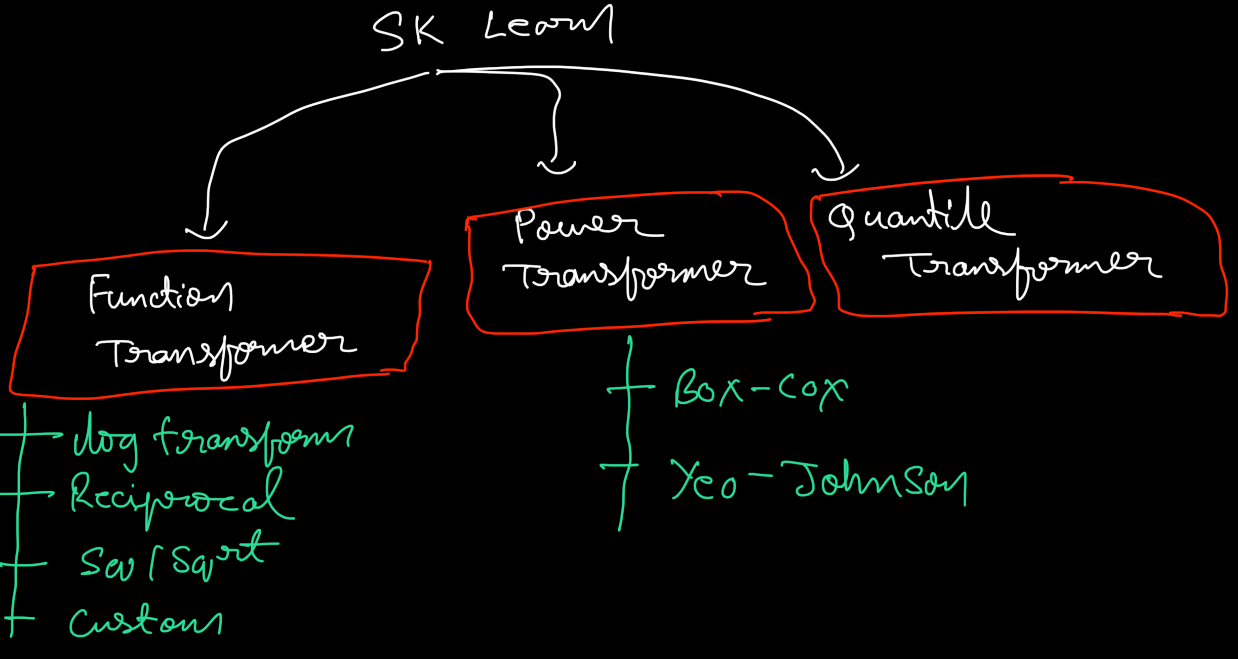
↳ Data distribution changes into Normal distribution



Kuch ML Algo Normal data Pe accha kaam karte hai

Transformation ka use kr ke hum data ko normal distribution में ला सकते हैं.

Note : Why Normal distribution is very important in Machine learning and Statistics domain ?



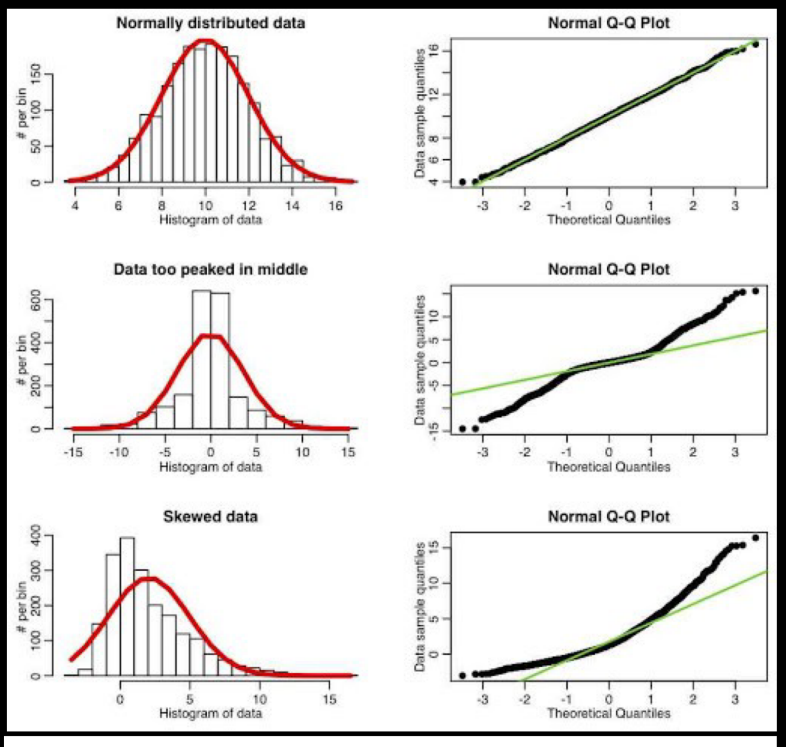
How to find if data is normal?

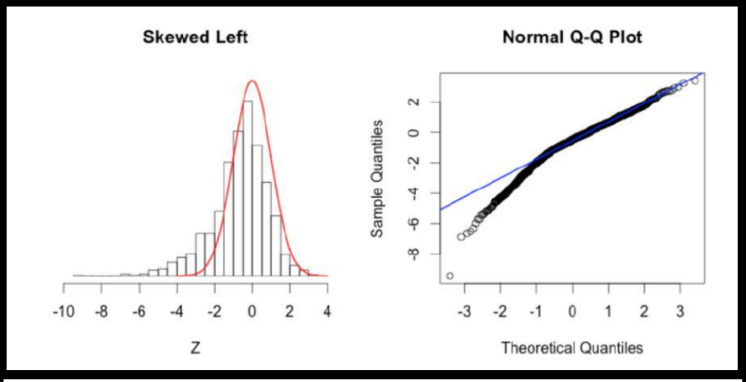
(i) use distribution plot (seaborn.distplot())

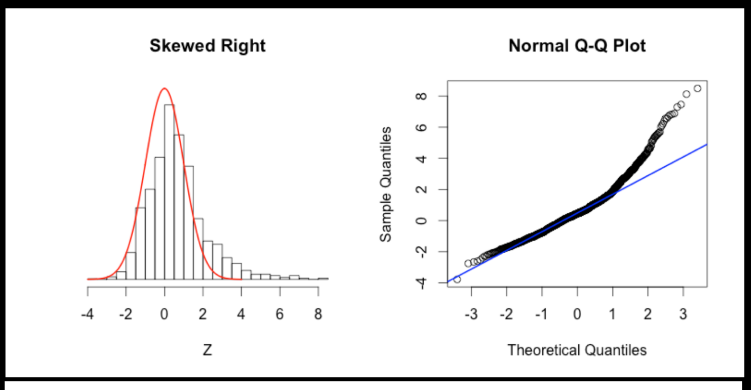
(ii) use Skew function (Pandas.skew())

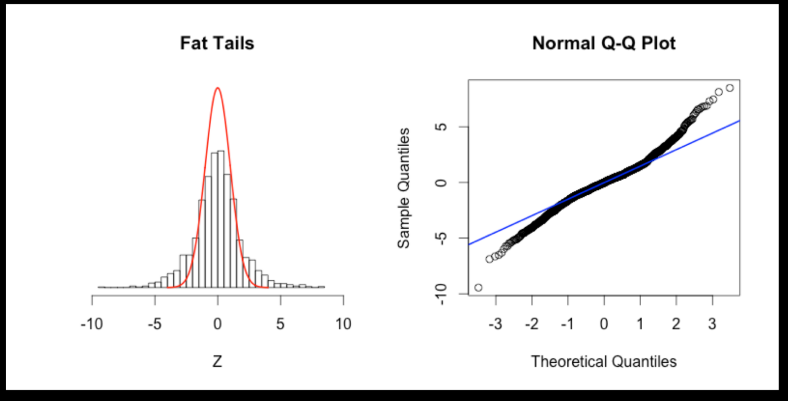
↳ Skewness should Zero

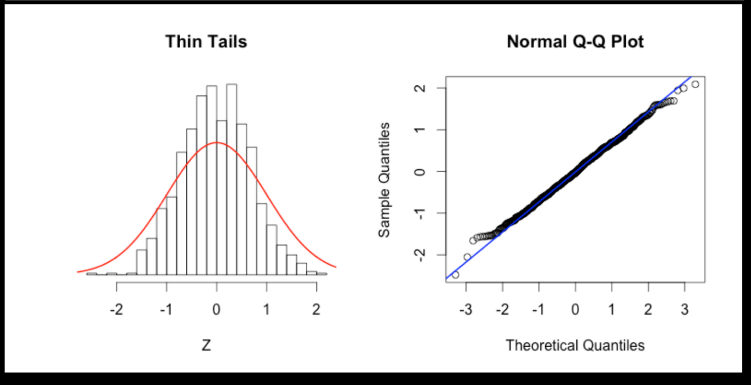
(iii) Q-Q Plot [Important in Statistics]











Log Transformation

↳ we use log

↳ not applicable in -ve data (we can not take log of -ve numbers)

↳ In right skewed data if we use log transformation then it brings data in centre.

↳ Bring every data in equivalent scale, so we start getting normal values and it start looking linear that's why linear model such linear Reg & logistic Reg start working better.

↳ log Convert additive Scale into multiplicative Scale.

Article on Log Transformation ⇓

https://medium.com/@kyawsawhtoon/log-transformation-purpose-and-interpretation-9444b4b049c9

Reciprocal Transformation

Smaller value → into Large value

Large value → into smaller value

Sq(x^2)

↳ for left Skewed data

Sqrt $\sqrt{x}$

* These transformation are like hit and trial, we don't know which work better in which Solution, So it is better to test all and check in which you are getting better result