M T W T F S S

Page No:

Date:

YOUVA

EXPT. NO.

NAME

(3) Intermediate steps should be transformers - All steps before the last one must have both .fit() and .transform() methods (like StandardScaler(), etc.

(4) Consistency - The same transformations are automatically applied to both training and test data, reducing the chances of data leakage.

(5) Parameter tuning - You can perform hyperparameter tuning directly on the entire pipelin, not just the model.

(6) Clean and resuable code - Pipelines make your code easier to maintain and ~~res~~ reuse for other datasets.

(7) Avoid data leakage - Since transformations are fitted only on the training set inside the pipelines, your model won't get biased from the test data.

✱ Function Transformer :- (Day - 30).

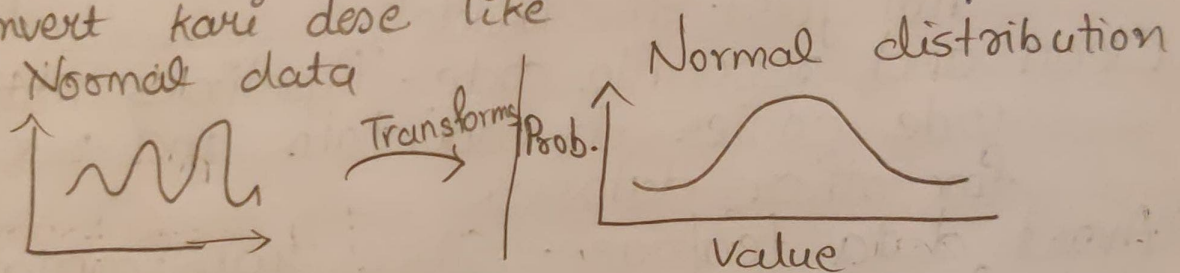In feature transformation:-

✱ Mathematical Transformation :-

"M.T. means applying a mathematical function to a variable in order to change its scale, spread, shape, or distribution. It helps make data more suitable for analysis and machine learning models."

⇒ Common Mathematical Trans. :- (Simple) (Complex):-

(1) Log Transformation:- | (1) Box - Cox
(2) Reciprocal Trans.:- | (2) Yeo - Johnson
(3) power (sq/sqrt):- | ₹

→ The main purpose to do this transformation is to tranf transform the data into the normal distribution form.

→ Etle keh aa transformation use etla mate thei keh like tamara pose always random ane distorted data hase to aa transformers. ene normal. distribution ma convert kari dese like

Normal data

Transform Prob.

Normal distribution

Value

→ It is mostly required for the algo. like
- linear regression   - Logistic regression.

Function Transformers:-

→ In sklearn library, there are 3 types of mostly used transformers' which are
  ↳ mathematical.
(1) function Transformers
(2) Power Transformers
(3) Quantile Transformers. (Not mostly used)

M T W T F S S
Page No.
Date
YOUVA
EXPT.
NO.
NAME

(1) function Transformers:-

$\left.\begin{array}{l} \text{L Log Trans.} \\ \text{L Reciprocal.} \\ \text{L sql sqrt} \\ \text{L custom.} \end{array}\right\}$ (Types of FT).

"It is a tool in data preprocessing that allows you to apply any mathematical function to transform a feature."

→ It is used when you want to apply a custom or specific transformation directly to your data.

→ function Transformer etle dataset par koi pan mathematical function (log, reciprocal, sql sqrt,) apply kaava no easy way. Aa transformer thi tame direct values per custom transformations pan kari sako, je data ne modeling mate vadhu suitable banave che.
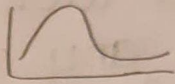
→ How to find if data is normally distributed?
(1) sns.distplot(). → seeing graph we can analysis.
(2) pd.skew() ⇒ If = 0 than. normally distributed.
 else (− or +) than is skewwed.
(3) QQ plot (most reliable) → Two lines (if overlapped
 perfectly than
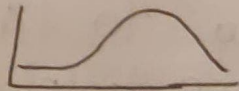 normally dist.)

Teacher's Signature:

⇒ Types of F.T :-

(1) log Transform :-

→ Apply the natural logarithm to the data.

⇒ Useful for highly skewed, positive data.

→ Eg :- $x' = \log(x)$.

→ Log transformation data ne compress kari ne skewness ghataadva madad kare che.

→ Used in right skewed data :- ⌒

(2) Reciprocal Transform :-

→ Takes the inverse of each value.

⇒ Helps when large values dominate the dataset.

→ Eg :- $x' = 1/x$.

→ Reciprocal ma value ne $1/x$ kari ne mota values nu impact ochhu kariye che.

(3) Square Transform :-

→ Square increases the impact of larger values.

⇒ Used for left skewed data :- ⌒

⇒ Square value ne vadhare highlight kare che.

⇒ Eg :- $x' = x^2$

(4) Square-root Transform :-

→ Square Root reduces moderate skewness.

→ Eg :- $x' = \sqrt{x}$.

→ Square root moderate skewness ne smooth kare che.