Feature Selection Techniques:-

Meaning:- reducing input variables for better model performance.

Reasons:-

→ Curse of dimensionality:-

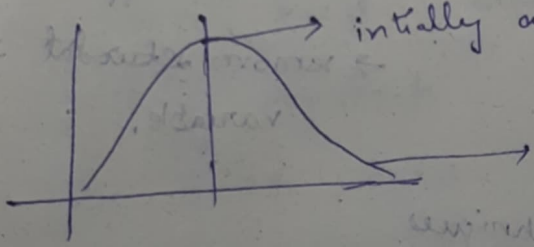
What is this?

When the number of input variables are high then model faces these problems,

→ model may learn from irrelevant data (overfitting)

→ slow learning of model / slow training.

→ visualization will be tough



initially as data is more model will perform good

but as time passes it learns irrelevant data (overfitting)

THIS PROBLEM CAN BE ADDRESSED IN TWO WAYS:-

- dimensionality reduction
  - ↓
  - in this no. of input variables will be reduced but feature transformation is possible
- feature selection:
  - ↓
  - no feature transformation features will be removed / added without change.

Lets look into feature selection:-

feature selection

Supervised

→ generally, target variable plays a imp role in supervised learning.

→ in this we will remove inputs which are irrelevant to target variable.

ex:- Sugar measure
heart measure → 3 inputs
bp measure

target variable :- diabetes

bp → not relevant.

Unsupervised

→ these type of models doesn't depend of target variable. it rather sees pattern in it.

→ so in this it check correlation b/n inputs and removes :-

Correlation :-

1 → more dependent

0 → no dependent

-1 → opposite

→ removes redundant variable.

→ Supervised :-

techniques

- Filter
- Wrapper
- Embedded.

Filter :-

When it is used :-

→ When dataset has many features and large.

→ The input variables are selected on the basis of relation

with target.

→ The selection takes place on the basis of certain statistics measure like :- correlation
variance
threshold etc...

drawback :-

→ not accurate as sometimes it may fail in input extraction if all the input values correlation or variance are same.

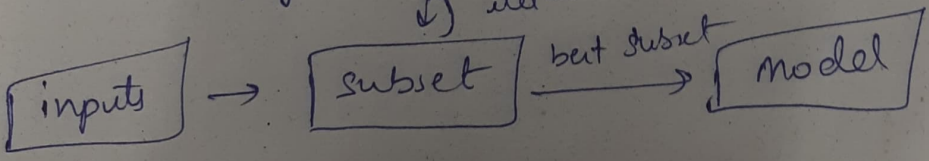
Wrapper :-

When it is used :-

→ dataset is too small and when you need accurate result.

→ Greedy approach.

→ This model takes different subsets of input features such a way that those features results in best performance model.



drawback :-

→ as it is iterative in nature it is expensive in time.

Embedded:-

↳ This is similar to upper given method as it extracts those features which contribute most to the training of a particular iteration.

↳ This uses penalization technique which is similar to reinforcement.

↳ Regularization methods:-

↳ Ridge Regression, LASSO etc...

Drawback:-

→ even though it is easy to implement it is more complex than filter technique.