

Session - 5.4

- Feature Selection -
 - feature \rightarrow i/p column
 - target \rightarrow o/p "
- We don't use all the features rather only use some. this is feature selection
- Why to do feature Selection -
 - 1) Curse of dimensionality -
 - ↳ In model only a certain number of feature will give optimum result.
 - If we increase # features then due to sparsity result good result will not come
 - 2) Computational complexity
 - ↳ reduce column reduce space & time complexity
 - 3) Interpretability -
 - ↳ if less columns then we can interpret easily
- Techniques/types of f.s
 - ↳ filter based technique
 - ↳ wrapper method
 - ↳ Embedded technique
 - ↳ hybrid technique

1) Filter based technique -

↳ we do some statistical measure in each column individually and filter the columns.

In this we have -

- Variance Threshold
- Correlation
- Anova
- ChiSquare
- Mutual Info

Step-1

	f_1	f_2	f_3	o/b
1	1	2	1	
2	2	1	2	
3	3	3	3	

→ Here we see that $f_1 \Delta f_3$ are duplicates
So we remove them

Step-2

1. Variance Threshold

↳ It is applied on two types of features

1) Constant

A	B	C	Here we see that B column
2	1	X	is constant so it has
7	1	N	no variance hence remove
2	1	N	it.

2) Quasi-constant feature

→ assume a feature in which ~~is~~ out of 1000 rows in 995 rows value is 1 and in remaining 5 rows value is 0. so variance is 0 so remove it.

- Step →
- decide threshold (like 0.1)
 - for all feature calculate threshold variance
 - drop if var. is less than threshold

→ There are some points to be considered -

- We only focus on individual column. Suppose f_1 has high var. So we kept it but what if it doesn't have any relfth with o/p and vice-versa means f_2 has low var. So we remove it but it has high relfth with o/p
- We don't see about relfth between f_1 & f_2 . what if f_1 has low var. but is selected to f_2
- Must do Scaling
- Challenging to find threshold.

→ Correlation (Pearson corr coeff)

↳ we find correlatth b/w i/p & o/p

↳ it is b/w -1 to 1

Strong inverse Strong positive
linear relfth linear relfth

→ 0 mean no linear relfth (may be non-linear)

↳ we can do -

$f_1 \rightarrow y$ } Find corr b/w f_1 & y

$f_2 \rightarrow y$ } if outside this

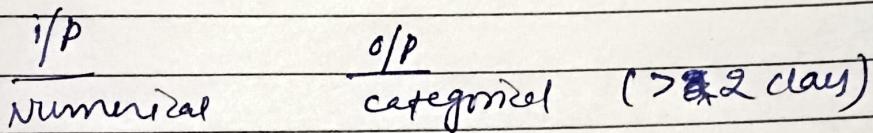
1 } $\rightarrow y$ } $\{0-0.3 \text{ or } 0.5\}$ then
remove feature

or we can find corr b/w f_1, f_2 etc and
if having more correl^h then keep one of them

Disadvantages -

- It's possible that there is non-linear relat^h
but we still drop it
- It doesn't capture complex relationship -
we only see relat^h b/w f_1, f_2, f_3 but what
if f_1, f_2, f_3 has relat^h?
- How to know threshold?
- Sensitive to outlier.

→ ANOVA - (Hypothesis test)



$f \rightarrow o/p$ (find relat^h b/w them)

if strong then keep it.

→ as strength of relat^h is find using
f - statics

Disadvantages -

- All groups that are formed should be normally distributed
- Variance of groups must be homogeneous
- Effect of outliers
- Doesn't take account in features relat^h

5) Chi-square test

↳ Already done

Advantage of filtered based feature selection

- Simplicity
- Speed
- Scalability
- pre-processing steps

• Disadvantage.

- don't consider interact^h b/w features
- model agnostic
- Threshold determination