Hello students myself Sneha Dessai

assistant professor in computer science

from Government College of Arts,

Science and Commerce quepem Goa.

Course title is introduction to data analytics

and I'm going to speak on feature

selection algorithms which comes

under the title features generation.

And feature selection.

So today I'm going to speak on the

topic of feature selection algorithms,

which are sequential forward

feature selection,

Sequential Backward feature selection,

combined approach, and embedded methods.

Under that decision, trees.

So at the end of this session,

students will be able to explain

various feature selection methods

that is forward selection,

backward elimination and combined approach.

And also they will be able

to explain embedded methods,

that is decision trees.

Basically there are three primary

methods of stepwise regression.

One is forward selection,

backward elimination and combined

approach which is bidirectional search.

So first one forward selection.

In forward selection,

you start with a regression

model with no features.

So empty set of features empty

set is taken and gradually

we add one one feature at a time.

We will add the features that improves

the model or sequential forward selection

in sequential forward selection.

First the best single features is

selected using some objective function.

Now using that best single

feature that will be paired with

all other remaining features.

And you identify which will

be the next best pair.

Once the best pair is identified.

You form the triplets of

features using other features.

And then identify the best triplet

so this procedure continues until a

predefined number of features are selected.

So this is the algorithm of

sequential forward feature selection.

Initially your set is empty.

You use find next algorithm to

find the feature and that feature

is added to the set S and you

remove that feature which is added

from the list of features.

So until your set satisfies the

major condition, that is the measure.

So this is the algorithm for sequential

forward features set generation.

But look at this example diagram, suppose if you have feature 12345.

At first stage I say that feature three

is the best feature among all others.

So I select feature three and

using feature 3I form pairs

with all other features.

So feature three and 3&1, 3&2, 3&4 and 3&5

Now out of this space I say that

three and one makes the best pair.

So I select that.

I make triplets using other features

SO 31
: 312 and 314 and 315.

So out of this triplets,

I say that three,

one and five is the best triplet.

I will stop at this point

because if I add more features,

the output of my model has degraded.

So I will stop with 315.

Or maybe I have said said that

only three features should

be included in the model,

so there should be certain stopping criteria.

So we have got we just rebound and

five each other best subset using

forward feature set generation.

This is the example here.

The result of sequential forward

feature selection for classification of

satellite image using create

features was drawn on the graph.

The X axis denotes the classification

accuracy and the Y axis denotes the

features added at each iteration.

The highest value accuracy value

is shown with the help of.

Top backward elimination

in backward elimination.

It is just the opposite of

your forward feature selection

in forward feature selection,

we start with the empty set and

we keep on adding the features.

The best features one at a time.

In backward elimination you start

with the regression model that

includes all the features and you

gradually remove one one feature and you

remove the worst feature at a time.

OK, so you stop removing the

features when removing the features.

Makes the selection criterion gets.

Worse, so sequential backward elimination.

First the objective function is

computed for all N features.

Then each feature is deleted one at a time.

And checked the output of the

model is checked now whichever.

Of feature.

Will be the worst if that is removed

then the model will give you a good output.

Now the worst feature if it is removed,

the output of the model will be good,

so that is eliminated next again

you train the objective function

using remaining n -- 1 features.

An other features are deleted one at a time.

So worst feature is discarded from

the subset with only n -- 2 features.

Now this procedure continues until.

Predefined number of features are left.

This is sequential.

Backward features are generation algorithm.

You initialize the set subset an you

keep on eliminating the features

from the subset. This is the example.

So initially we go with features 12345.

We see that feature one is the most feature,

so we eliminate that and we get only

features 2-3 and four and five.

The out of this we say that feature

four is if it is eliminated then

our model works properly.

So we remove feature four and we

go with only features 2-3 and five.

So this is the.

A stopping criteria means we will if

we try to eliminate more features,

output will become worse.

So what we do,

we retain all these features and

this is our final feature subset

which feature 2-3 and five.

So this is the example of

sequential backward selection.

So graph is plotted.

Again,

classification accuracy is on the

X axis and features removed at each

iteration on the Y axis and the

star denotes the highest accuracy.

Combined approach,

which is bidirectional search.

We apply both sequential forward

selection and sequential backward

selection simultaneously.

Now sequential forward selection.

We know that it is performed on the

empty set whereas sequential backward

selection is performed on the full set.

Now to guarantee that both forward

and backward selection converge

to the same solution,

we have to see that features

which are already selected by

the forward feature selection.

I'm not removed by backward search,

whereas features which are already

removed by backward search are

not added by forward search.

So only if we check on this two

condition then we will get the

correct solution for the model.

So what are the limitations of

sequential forward selection and

sequential backward selection?

The main limitation of sequential

forward selection is that it is unable

to remove features that become non

useful after addition of other features.

The main limitation of sequential

backward selection is its inability

to revaluate the usefulness of a

feature after it has been discovered.

Next is feature selection.

Embedded methods now in embedded

methods you do not use any filter

because the information gain method is

doing your feature selection process.

So one example of embedded method

is decision trees.

So decision trees have an initiative

appeal because outside the context

of data science in our everyday

lives we can think of breaking big

decision down into series of questions.

So decision tree.

This is the decision tree for

college students.

The decision tree is actually dependent

on bunch of factors whether or not

there are any parties or deadlines.

How lazy is the student filling

and what they care about most.

They care about parties, right?

So here is there a party they will

go for the party if there's no party,

then you check whether there is a deadline.

If the deadline is urgent, they will study,

otherwise they will go to pub.

If deadline is Neo to check whether

the student is lazy.

If student is lazy, what it will they want?

The student will do.

Students will watch TV.

Otherwise, students will study.

So this is the decision tree.

Now, this is another decision

tree for chasing Dragons.

We have to check whether the user will

come back to use your chasing dragon F.

So first we check whether the

user has fill out the app profile.

If yes, did users claim more than 10 Dragons?

Yes,

then a prediction is user will

come back if user has not played

more than 10 Dragons then did user

invite at least one friend to join.

If yes,

then prediction is user will come back

otherwise user will not come back.

If you go on the other side.

If user has not filled out the profile,

you check whether the user has

come back more than once in.

Firstly, if yes,

prediction is user will come back.

Otherwise prediction is user

will not come back.

So decision trees they take as input

an object or situation described by set

of attributes and returns a decision.

So this decision is the predicted

output for the input in decision trees.

All non leaf nodes lead to the

partial decision and assists in

moving towards the leaf node and leaf

nodes depict the decision about a

character having attributes falling

on the path from the root node.

So what are the limitations?

The tree memorizes the observation

but does not extract any pattern

from the examples.

This limits the capability of

learning algorithm in that the

observations do not extrapolate

to examples it has not seen.

So the decision tree algorithm is first step.

You build your decision tree iteratively,

starting at the root.

Then you need an algorithm to

decide which attribute to split.

An example which node should be next

one to identify next you choose the

attribute in order to maximize the

information gain and last step is

you keep going until all the points

at the end are in the same class or

you end up with no features left.

In this case, you take the majority vote.

If no features are left.

these are my references, thank you.