# Bagging.R

patriciamaya

2020-09-29

```
require(data.table)
```

```
## Loading required package: data.table
```

```
library(rpart)
```

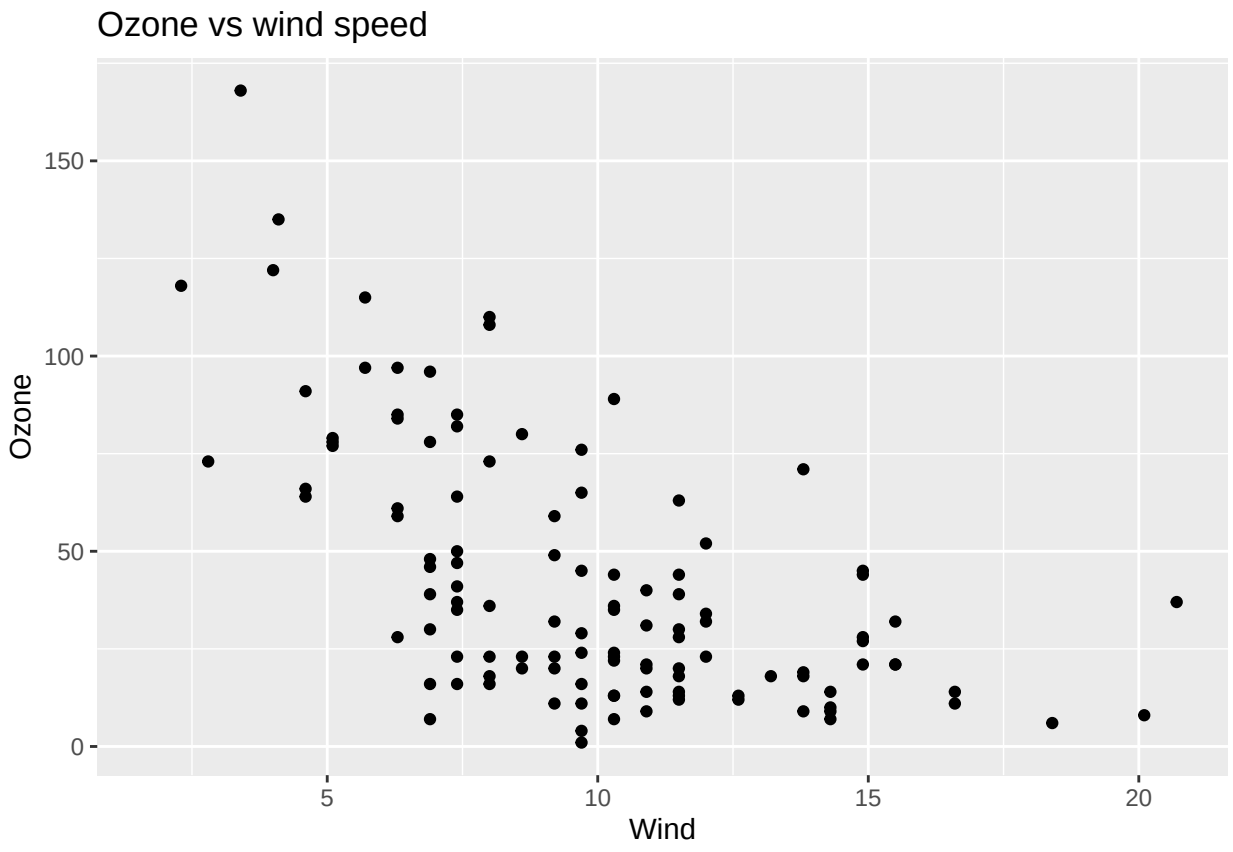
```
## Warning: package 'rpart' was built under R version 4.0.2
```

```
require(ggplot2)
```

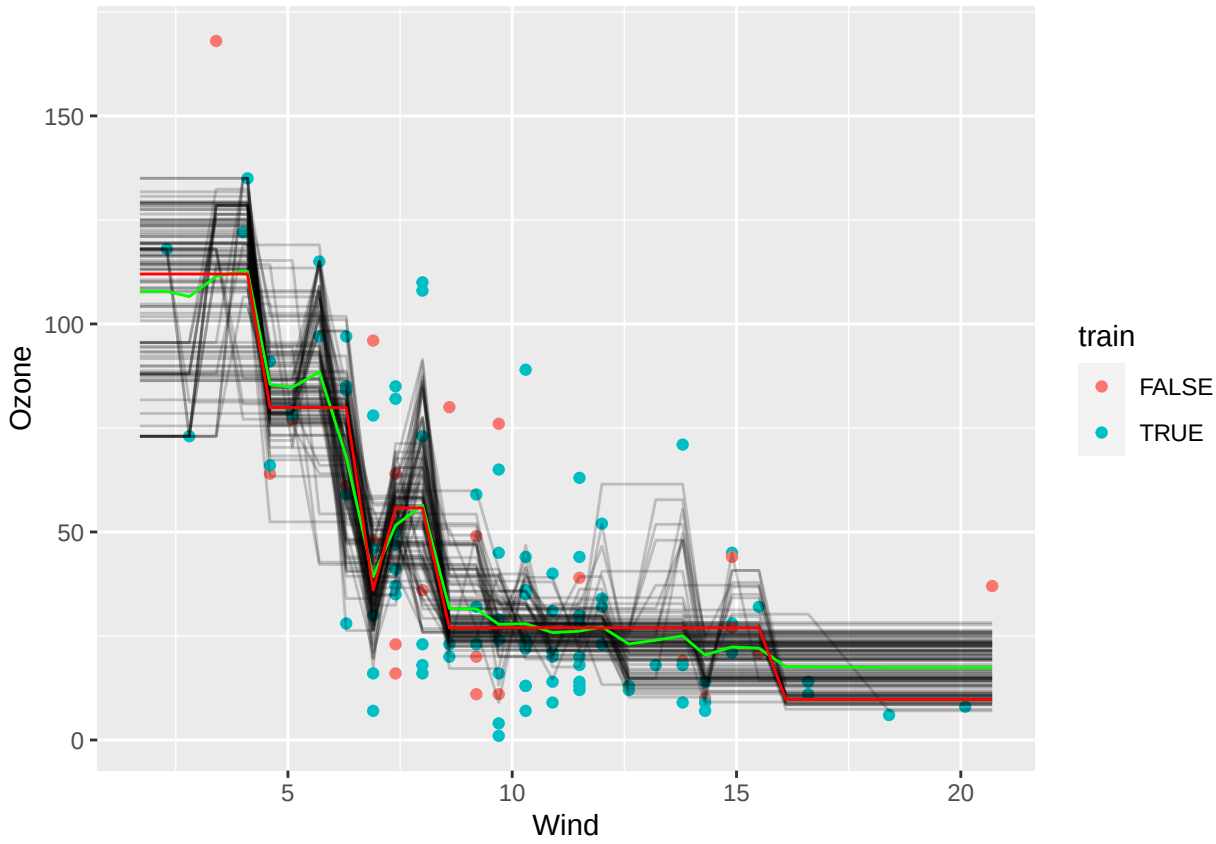
```
## Loading required package: ggplot2
```

```
set.seed(456)
##Reading data
bagging_data=data.table(airquality)
ggplot(bagging_data,aes(Wind,Ozone))+geom_point()+ggtitle("Ozone vs wind speed")
```

```
## Warning: Removed 37 rows containing missing values (geom_point).
```



```
data_test=na.omit(bagging_data[,.(Ozone,Wind)])
##Training data
train_index=sample.int(nrow(data_test),size=round(nrow(data_test)*0.8),replace = F)
data_test[train_index,train:=TRUE][-train_index,train:=FALSE]
##Model without bagging
no_bag_model=rpart(Ozone~Wind,data_test[train_index],control=rpart.control(minsplit=6))
result_no_bag=predict(no_bag_model,bagging_data)
##Training of the bagged model
n_model=100
bagged_models=list()
for (i in 1:n_model)
{
  new_sample=sample(train_index,size=length(train_index),replace=T)
  bagged_models=c(bagged_models,list(rpart(Ozone~Wind,data_test[new_sample],control=rpart.control(minspl
}
##Getting estimate from the bagged model
bagged_result=NULL
i=0
for (from_bag_model in bagged_models)
{
  if (is.null(bagged_result))
    bagged_result=predict(from_bag_model,bagging_data)
  else
    bagged_result=(i*bagged_result+predict(from_bag_model,bagging_data))/(i+1)
  i=i+1
}
##Plot
require(ggplot2)
gg=ggplot(data_test,aes(Wind,Ozone))+geom_point(aes(color=train))
for (tree_model in bagged_models[1:100])
{
  prediction=predict(tree_model,bagging_data)
  data_plot=data.table(Wind=bagging_data$Wind,Ozone=prediction)
  gg=gg+geom_line(data=data_plot[order(Wind)],aes(x=Wind,y=Ozone),alpha=0.2)
}
data_bagged=data.table(Wind=bagging_data$Wind,Ozone=bagged_result)
gg=gg+geom_line(data=data_bagged[order(Wind)],aes(x=Wind,y=Ozone),color='green')
data_no_bag=data.table(Wind=bagging_data$Wind,Ozone=result_no_bag)
gg=gg+geom_line(data=data_no_bag[order(Wind)],aes(x=Wind,y=Ozone),color='red')
gg
```


150
100
50
0
Ozone
5
10
15
20
Wind
train
FALSE
TRUE