Introduction to Naive Bayes (Udemy Course)

Heaven Klair

6/6/2022

Importing the dataset

```
dataset = read.csv('Social_Network_Ads.csv')
Age = dataset$Age
EstimatedSalary = dataset$EstimatedSalary
Purchased = dataset$Purchased
```

Splitting the dataset into the Training set and Test set

You can also embed plots, for example:

```
library(caTools)
set.seed(123)
split = sample.split(dataset$Purchased, SplitRatio = 0.75)
training_set = subset(dataset, split == TRUE)
test_set = subset(dataset, split == FALSE)
```

Feature Scaling

```
training_set[-3] = scale(training_set[-3])
test_set[-3] = scale(test_set[-3])
```

Fitting classifier to the Training set

```
library(e1071)
classifier = naiveBayes(x = training_set[-3],
                        y = training_set$Purchased)
```

Predicting the Test set results

```
y_pred = predict(classifier, newdata = test_set[-3])
```

Making the Confusion Matrix

```
cm = table(test_set[, 3], y_pred)
cm
```

```
##    y_pred
##      0  1
##   0 57  7
```

```
##    1  7 29
```

Visualising the Training set results

```
#
```

```
# install.packages('tidyverse')
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.6      v dplyr   1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# expand.grid('Age' = seq(min(set[, 1]) - 1, max(set[, 1]) + 1, by = 0.01),
# 'EstimatedSalary' = seq(min(set[, 2]) - 1, max(set[, 2]) + 1, by = 0.01))%>% mutate(prob_set=predict(
# y_grid = ifelse(prob_set > 0.5, 1, 0))%>%
# ggplot() +
# geom_point(aes(x=Age, y=EstimatedSalary, color=y_grid)) +
# geom_point(data=training_set, aes(x=Age, y=EstimatedSalary,colour=as.numeric(Purchased)))
```

Visualising the Test set results