



Erick Vargas

Machine Learning

Machine Learning

Contents

1 Lesson 1 - Welcome to machine learning	2
1.1 Prerequisites	2
1.2 Introduction	2
2 Lesson-2 What is Machine Learning?	3
2.1 What is machine learning?	3
2.2 Decision trees	3
2.3 Naive bayes	4

Lesson 1 - Welcome to machine learning

1.1 Prerequisites

To this course we need to know some things

- Python
- Probability and statistics
- Linear algebra

1.2 Introduction

Machine learning have a lot of useful applications, some of them are:

- Voice recognition
- Spam detection
- Fraud detection
- Stock market
- Teaching a computer how to play chess
- Self-driving cars

Lesson-2 What is Machine Learning?

2.1 What is machine learning?

To make a really easy definition let's think about this: in the world exist humans and computers, we already know that humans get knowledge from past experiences in the other hand computers need to be programmed to know what to do so we need to made the question: ¿Computers can also get knowledge from past experiences too? and this is the answer that machine learning can give us. Obviously for a computer "past experiences" are just data, so we go to learn how to use this data to develop machine learning.

2.2 Decision trees

Let's start with a simple example, imagine that we need to recommend applications from the google play store to each user.

Gender	Age	App
F	15	Pokemon go
F	25	Whatsapp
M	32	Snap chat
F	40	Whatsapp
M	12	Pokemon go
M	14	Pokemon go

Consider the table above and answer Between **gender** and **age**, which one seems more decisive for predicting what app the users will download? the answer is very simple, is the age, as we can see we get more information about the preferences between all the applications that we have in the table. But let's do this analysis step by step, let's split the table first using the gender:

Gender	Age	App
F	15	Pokemon go
F	25	Whatsapp
M	32	Snap chat
F	40	Whatsapp
M	12	Pokemon go
M	14	Pokemon go

As we can see female users have downloaded Pokemon go and WhatsApp in the other hand male users have downloaded Snap Chat and Pokemon Go. So we can get good sets using the gender.

Now we go to make the same analysis but now we go to use the age, let's consider the users that have 20 or less age as one set and more of 20 as the other one.

Gender	Age	App
F	15	Pokemon go
F	25	Whatsapp
M	32	Snap chat
F	40	Whatsapp
M	12	Pokemon go
M	14	Pokemon go

As we can see we get better results using the age. For example the people under 20 years old use to download pokemon go, does not matter if there are male or female. But it is a little bit different with whatsapp and snap chat, in this case if the user is female the application that they used to use is whatsapp otherwise is snap chat. We can see this using the next diagram.

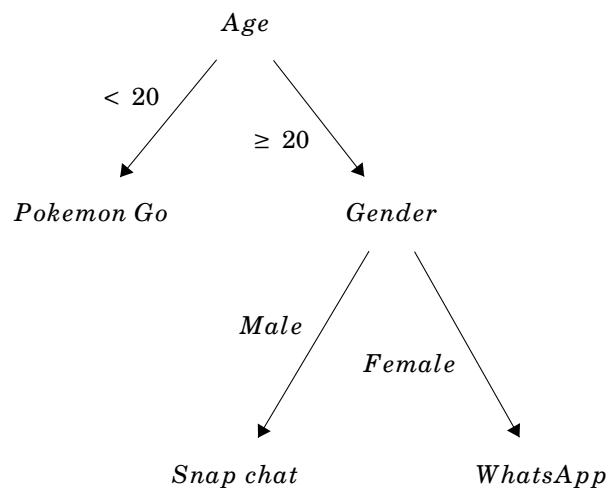


Figure 2.1: Decision tree

2.3 Naive bayes

To understand this we go to imagine that we need to build a detecting spam e-mails. To build this let's suppose that we have a hundred of emails that are data, and we already looked that. Now we suppose that twenty five emails are spam. So now let's try to think of features that a spam email may be likely to display. Let's suppose that one feature of this emails are the word cheap but five of these emails are not really spam emails. So how can we answer the next question, If an e-mail contains the word "cheap", what is the probability of being spam? We know that we have 25 candidates to being spam, but only 5 emails are really spam so just the 20% so just the 80% are really spam. It is the naive bayes algorithm.