Are you born a drug user?

Christophe Jefferies
Imperial College London

Introduction

Do background, personality, and use of legal substances affect people's use of illegal substances? This poster seeks to answer this question.
Why is this question important?

- Predicting illegal drug use early can lead to preventing it later
- Measuring how a person's characteristics affect probability of drug use is also of great to researchers, in particular in psychology

The dataset

- Over 1800 people
- Detailed information for each person about background, personality, and use of legal and illegal substances

The following shows how many different illegal drugs people in the dataset had ever taken.

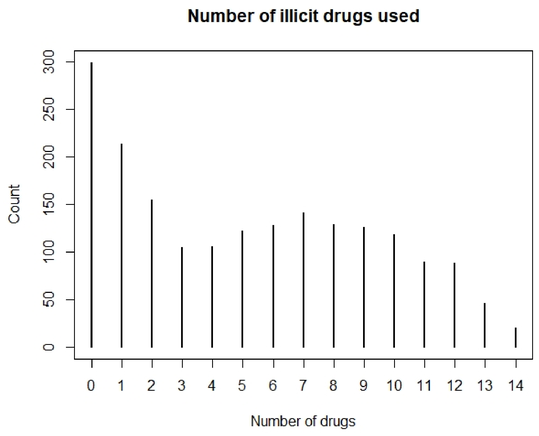

Figure 1: How many people reported using different numbers of illegal drugs

We seek to predict this statistic from a person's other characteristics.

Data visualisation

- This plot places two points close together if their background, personality, and legal substance use are similar (using a method called 'TSNE').
- The colours are then assigned to show number of different drugs ever used.

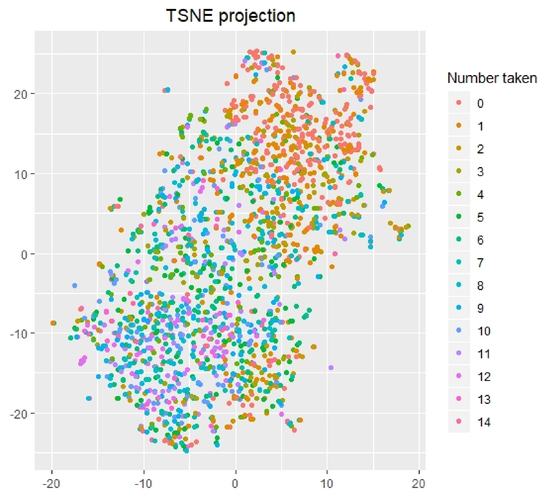

Figure 2: Data projection coloured by number of illegal drugs taken (to label the axes makes no sense for a TSNE projection)

- We see that similarly-coloured points are generally close together.
- There are clear patches of colour for very high and very low illegal substance use.
- This means we should be able to predict variety of drug use based on the other characteristics.

Approach and analysis

- We will use a logistic regression model to classify people's drug use according to their other characteristics
- The model takes in data like the following:

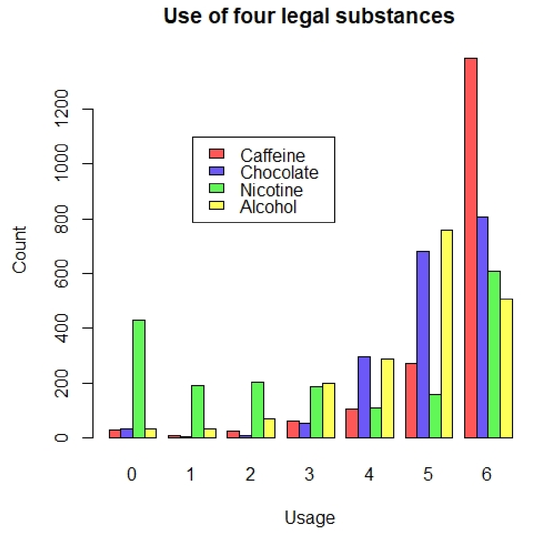

Figure 3: Reported use of four legal substances

- The model then tells us how much each characteristic affects the outcome.
- According to this, the most significant outcomes are:

1. Gender: male
2. Age range: 24-35 years
3. Personality trait: open to experiences
4. Country: America

Deduction

Our logistic regression analysis suggests: the best predictors of how many different illegal drugs someone has taken are being male, 25-34 years old, American, and being open to experiences.

Test quality

- We can assess the quality of our classifier using ROC analysis.
- This tracks how well our model performs as we vary the cutoff point where someone is predicted to have used many drugs

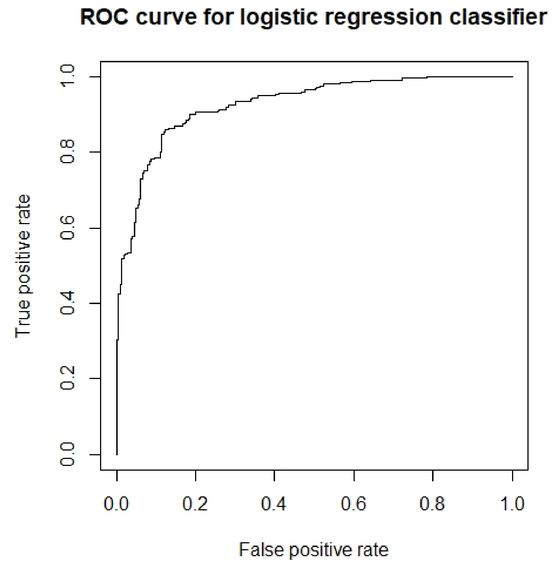

Figure 4: ROC analysis of the logistic regression classifier

A good measure of an ROC analysis is the area under the curve, which here is 0.927, a very good score.

Summary and discussion

In summary, characteristics such as background, personality, and use of legal substances are very good predictors of how many different drugs someone has used.
In particular, certain features are more telling than others, and we have found these most significant characteristics.
With a larger and more varied dataset, we could make similar conclusions; a future analysis might consider political opinion, fitness/weight, and so on.
The contents of this work and the associated code are my own unless otherwise stated