A Recommendation System for Movies

Joe King

12 June 2020

Introduction

The purpose of this project is to provide a machine learning model for a movie recommendation system. The data on which this model is produced is the Movielens 10M available from GroupLens research lab

The Movielens 10M dataset comprises around 10 million ratings categorised by user and movie. Approximately 10% of the data is reservered as a validation set, with the remainder being used to train the model.

The metric for measuring the quality of the model produced is the root mean square error (**RMSE**) represented by the equation

$$\textbf{RMSE} = \sqrt{(\tfrac{1}{\mathbf{N}}) \sum_{u,i} (\hat{y}_{u,i} - y_{u,i})^2}$$

where $y_{u,i}$ is the rating for movie i by user u, $\hat{y}_{u,i}$ is the corresponding prediction produced by the model, and N is is the number of user/movie combinations in the validation dataset.

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

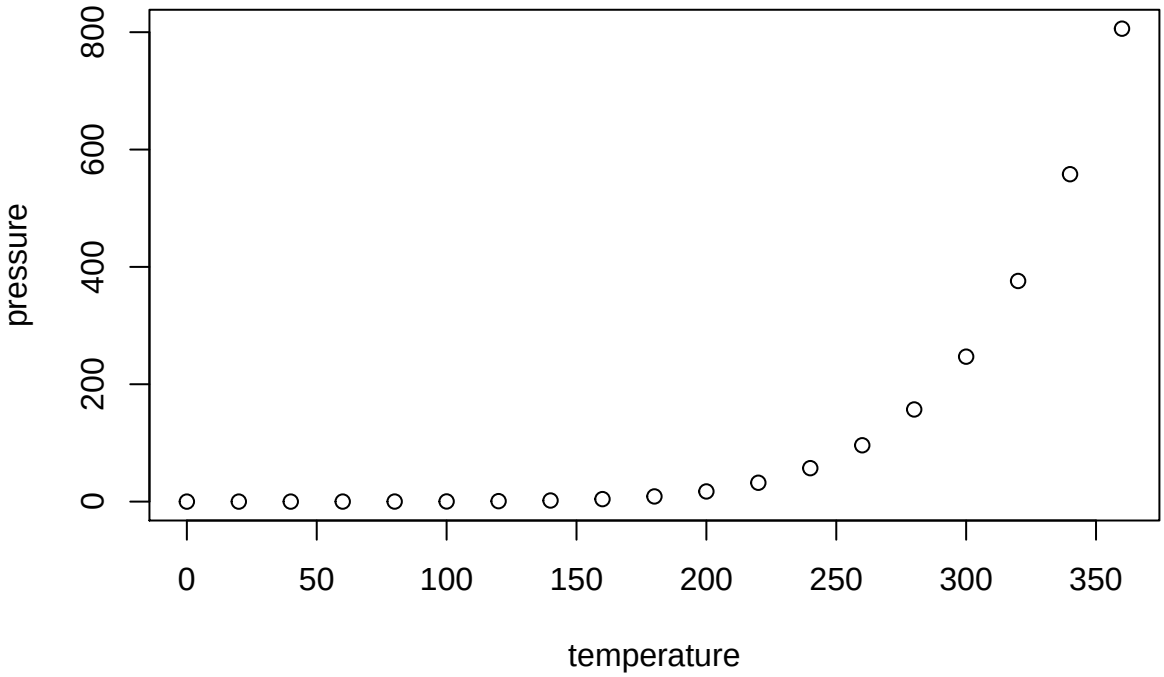
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.