

Capstone MovieLens Report

Azamat Kurbanaev

2025-02-03

Contents

Introduction / Overview / Executive Summary	1
Methods / Analysis	2
Conclusion	2

Introduction / Overview / Executive Summary

The goal of the project is to build a Recommendation System using a [10M version of the MovieLens dataset](#). Following the [Netflix Grand Prize Contest](#) requirements, we will evaluate the RMSE score defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i,j}^N (y_{i,j} - \hat{y}_{i,j})^2}$$

with N being the number of user/movie combinations for which we make predictions and the sum occurring over all these combinations[1].

Our goal is to achieve a value of less than 0.86490 (compare with the *Netflix Grand Prize* requirement: of at least 0.8563[2]).

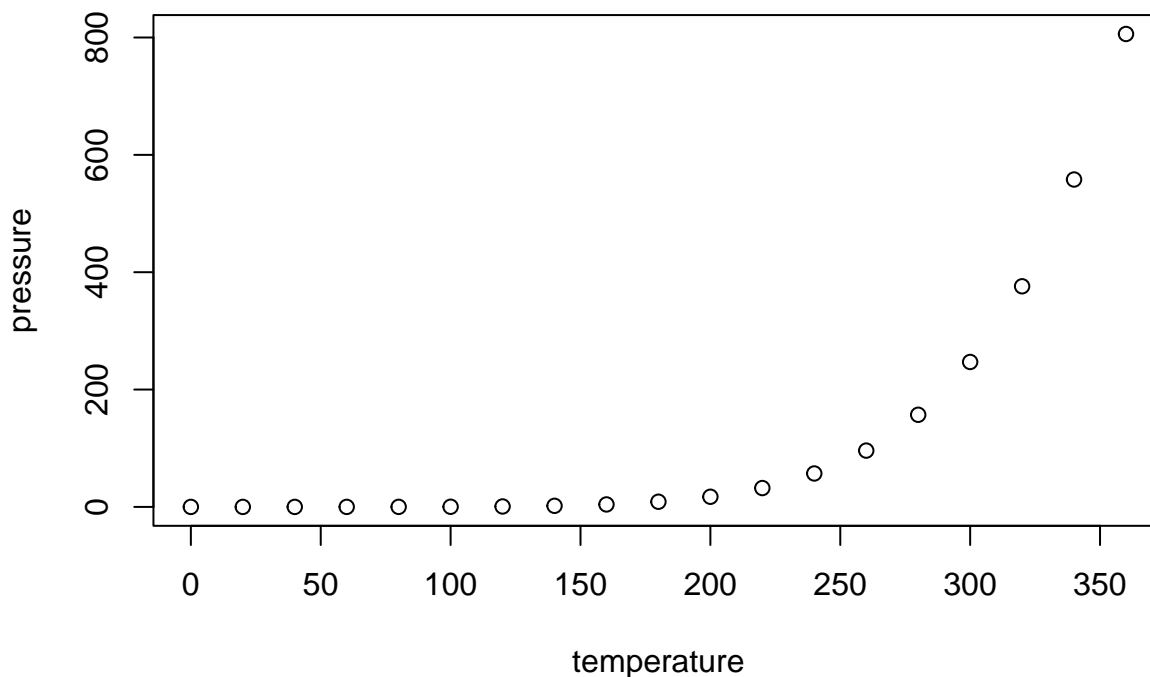
Dataset

```
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
```

Methods / Analysis

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.

Conclusion

Hello Conclusion!

This is a great conclusion, isn't it?!!

References

- [1] Rafael A. Irizarry. *Introduction to Data Science, Part II, Section 23.2: Loss function. Statistics and Prediction Algorithms Through Case Studies*. Dec. 27, 2024. URL: <https://rafalab.dfci.harvard.edu/dsbook-part-2/highdim/regularization.html#sec-netflix-loss-function> (visited on 12/27/2024) (cit. on p. 1).
- [2] Robert M. Bell Andreas Toscher Michael Jahrer. *The BigChaos Solution to the Netflix Grand Prize. commendo research & consulting*. Sept. 5, 2009. URL: https://www.asc.ohio-state.edu/statistics/statgen/joul_aut2009/BigChaos.pdf (visited on 09/05/2009) (cit. on p. 1).