

ML lab 2

Task 1:

Why don't we use Boston data set?

The Boston housing dataset is small, especially in today's age of big data. But there was a time where neatly collected and labeled data was extremely hard to access, so a publicly available dataset like this was very valuable to researchers. And although we now have things like Kaggle and open government initiatives which give us plenty of datasets to choose from, this one is a staple to machine learning practice as chocolate is to a break-up.

Each of the 506 rows in the dataset describes a Boston suburb or town, and it has 14 columns with information such as average number of rooms per dwelling, pupil-teacher ratio, and per capita crime rate. The last row describes the median price of owner-occupied homes (this leaves out homes that are rented out), and it's usually the row that we are trying to predict when we use it for regression tasks