

Machine Learning Exercise 0

Palle Morris e1160346
Elmenreich Jan e01526208
Holzberger Fabian e11921655

March 22, 2020

1 Introduction

Two datasets are analyzed, one for Classification and a second one for regression. The Datasets were chosen such that they have different characteristics. The Characteristics and more information about the datasets is listed in the table below.

Characteristic	Hepatitis	House sales
Data Type	Multivariate	Multivariate
Attribute Type	Integer,Real	Integer, String, Real
Associated Tasks	Classification	Regression
Number of instances	155	21436
Number of Attributes	19	20
Missing Values	Yes	No

2 Hepatitis Dataset

3 House sales Dataset

This dataset contains 20 attributes + `id`, which we will not count as an attribute. Our target attribute will be, as so often the `price`

- **Price Real** (target)
Price of each home sold
- **Date String**
Date of the home sale
- **Bedrooms Integer**
Number of bedrooms
- **Bathrooms Real**
Number of Bathrooms
- **Sqft_living Integer**
Square footage of the apartments interior living space
- **Sqft_lot Integer**
Square footage of the land space
- **Floors Real**
Number of floors
- **Waterfront Integer**
A dummy variable for whether the apartment was overlooking the waterfront or not
- **View Integer**
An index from 0 to 4 of how good the view of the property was
- **Condition Integer**
An index from 1 to 5 on the condition of the apartment
- **Grade Integer**
An index from 1 to 13, where 1–3 falls short of the building construction and design, 7 has an average level of construction and design, and 11–13 have a high quality level of construction and design
- **Sqft_Above Integer**
The square footage of the interior housing space that is above ground level
- **Sqft_basement Integer**
The square footage of the interior housing space that is below ground level
- **Yr_built Integer**
The year the house was initially built
- **Yr_renovated Integer**
The year of the house's last renovation
- **Zipcode Integer**
What zipcode area the house is in
- **Lat Real**
Latitude
- **Long Real**
Longitude
- **Sqft_living15 Integer**
The square footage of interior housing living space for the nearest 15 neighbors
- **Sqft_lot15 Integer**
The square footage of the land lots of the nearest 15 neighbors

In picture 1 we can see the distribution of the target and other attributes.

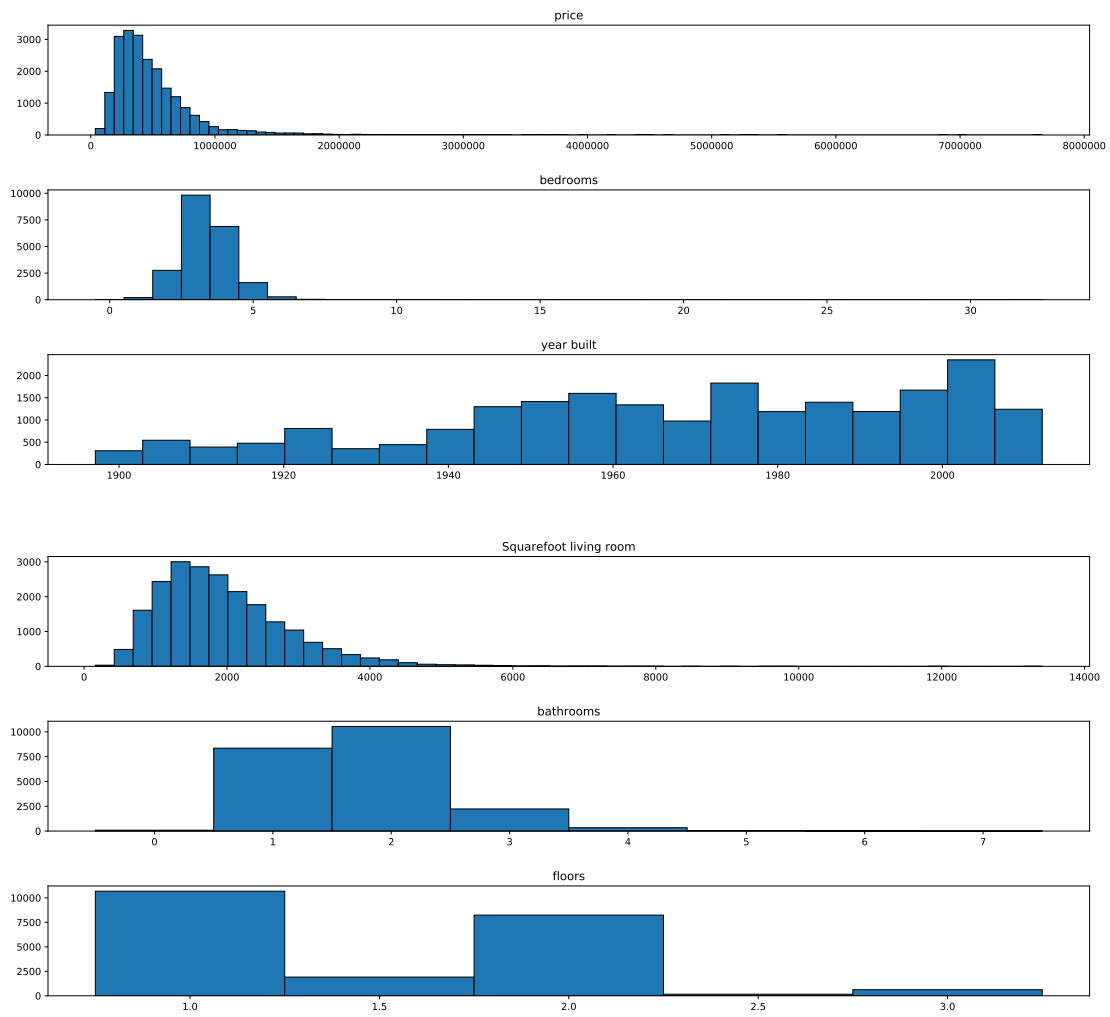


Figure 1: Histogramms of attributes