Linear Regression for Prediction

Machine Learning - Section 3.1.1

Marc Omar Haddad

Published: 9 February, 2020

Updated: 10 February, 2020

Linear Regression can be considered to be a form of Machine Learning. Although it is too rigid to be useful in general, it can be very effective in certain cases. It is also a baseline approach to Machine Learning in that it is often used if more complex methods are impractical.

Linking Linear Regression and Machine Learning

We can use Galton's data set to exhibit the link between Linear Regression and Machine Learning.

```
library(HistData)

galton_heights = GaltonFamilies %>%
  filter(childNum == 1 & gender == "male") %>%
  select(father, childHeight) %>%
  rename(son = childHeight)
```

Our task is to build a Machine Learning algorithm that predicts the `son`'s height Y using the `father`'s height X.

First, we generate our `test_set` and `train_set`.

```
y = galton_heights$son
test_index = createDataPartition(y, times = 1, p = 0.5, list = FALSE)

train_set = galton_heights %>% slice(-test_index)
test_set = galton_heights %>% slice(test_index)
```

To see if our eventual algorithm performs better than merely guessing, we create an algorithm that estimates `son` by simply finding the average of all `son` heights in our `train_set`.