Multivariate Regression on Boston House Prices

Notebook imports

```
%matplotlib inline import numpy as np import pandas as pd import matplotlib.pyplot as plt import seaborn as sns import sklearn.linear_model import sklearn.metrics
```

```
import sklearn.model_selection
```

```
import sklearn.preprocessing
```

```
import sklearn.ensemble
```

```
import sklearn.tree
```

```
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor
```

```
# Ignore warnings from old versions of scikit-learn
import warnings
```

```
warnings.filterwarnings('ignore')
```

```
%matplotlib inline
```

Gather Data

```
!wget -q -O boston_housing.h5 load_boston(n=
```

```
# Keras specific
```

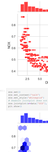
```
# If using TensorFlow backend.
```

```
np.random.seed(1337)
```

```
from keras.models
```

```
Using TensorFlow backend.
```

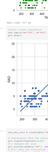
```
[nltk_data] Downloading package stopwords. [nltk_data] Unzipping corpus/stopwords.zip.zipfile. [nltk_data] Downloading package punkt. [nltk_data] Unzipping tokenizers/punkt.zipfile. [nltk_data] Downloading package wordnet. [nltk_data] Unzipping corpus/wordnet.zipfile. [nltk_data] Downloading package averaged_perceptron_tagger. [nltk_data] Unzipping taggers/averaged_perceptron_tagger.zipfile. [nltk_data] Downloading package maxent_ne_chunker. [nltk_data] Unzipping chunkers/maxent_ne_chunker.zipfile. [nltk_data] Downloading package words. [nltk_data] Unzipping corpus/words.zipfile. [nltk_data] Downloading package maxent_treebank_pos_tag. [nltk_data] Unzipping taggers/maxent_treebank_pos_tag.zipfile. [nltk_data] Downloading package conll2000. [nltk_data] Unzipping corpus/conll2000.zipfile. [nltk_data] Downloading package brown. [nltk_data] Unzipping corpus/brown.zipfile. [nltk_data] Downloading package tagsets. [nltk_data] Unzipping help/tagsets.zipfile. [nltk_data] Downloading package stopwords. [nltk_data] Unzipping corpus/stopwords
```




```
ggplot(data=Boston)+
  geom_point(aes(x=CRIM,y=Price))+
  geom_smooth(aes(x=CRIM,y=Price))
# Scatterplot visualizes data for crime rate against the median value of owner-occupied homes
# Geom_smooth adds a smooth line to the plot, helping to identify the trend.
# The plot shows a negative correlation, indicating that higher crime rates tend to be associated with lower property values.
```

Benchmarking

```
library
library(ggplot2)
ggplot(data=Boston)+
  geom_point(aes(x=CRIM,y=Price))+ geom_smooth(aes(x=CRIM,y=Price))+ ggtitle("Price vs CRIM")+ xlab("CRIM")+ ylab("Price")+ theme(plot.title=element_text(size=20,hjust=0.5))
# Enhanced version with proper labels and formatting
```

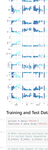



```
Boston.lm <- lm(Price~CRIM)
summary(Boston.lm)
# Fit linear model and examine results
ggplot(data=Boston)+
  geom_point(aes(x=CRIM,y=Price))+ geom_abline(intercept=24.03,slope=-0.42)
# Add regression line with specified intercept and slope
```

```
log_dis_lm = lm(Price~log(DIS))
# Create model using log transformation of distance
# Transformation helps with non-linear relationships
ggplot(data=Boston)+
  geom_point(aes(x=log(DIS),y=Price))+ geom_abline(intercept=24.03,slope=-0.42)
# Visualize the transformed relationship
```

```
ggplot(data=Boston)+
  geom_point(aes(x=log(DIS),y=Price))+ geom_smooth(aes(x=log(DIS),y=Price))
# Add smooth line to show pattern
```

```
ggplot(data=Boston)+
  geom_point(aes(x=log(DIS),y=Price))+ geom_smooth(aes(x=log(DIS),y=Price))+ ggtitle("Price vs log(DIS)")+ xlab("log(DIS)")+ ylab("Price")+ theme(plot.title=element_text(size=20,hjust=0.5))
# Enhanced version with proper labels and formatting
```




```
Boston.lm = lm(Price~CRIM+log(DIS))
summary(Boston.lm)
# Multiple regression with crime rate and log distance
ggplot(data=Boston)+
  geom_point(aligned(x=CRIM,y=Price))+ geom_abline(intercept=24.03,slope=-0.42)
# Attempt to visualize multiple regression
```

```
ggplot(data=Boston)+
  geom_point(aligned(x=CRIM,y=Price))+ geom_smooth(aligned(x=CRIM,y=Price))
# Add smooth line to show pattern
```

```
ggplot(data=Boston)+
  geom_point(aligned(x=CRIM,y=Price))+ geom_smooth(aligned(x=CRIM,y=Price))+ ggtitle("Price vs CRIM")+ xlab("CRIM")+ ylab("Price")+ theme(plot.title=element_text(size=20,hjust=0.5))
# Enhanced version with proper labels and formatting
```

Price vs CRIM