Linear Regression

In this section, you'll use linear regression to predict life expectancy from body mass index (BMI). Before you do that, let's go over the tools required to build this model.

For your linear regression model, you'll be using scikit-learn's `LinearRegression` class. This class provides the function `fit()` to fit the model to your data.

```
>>> from sklearn.linear_model import LinearRegression
>>> model = LinearRegression()
>>> model.fit(x_values, y_values)
```

In the example above, the `model` variable is a linear regression model that has been fitted to the data `x_values` and `y_values` . Fitting the model means finding the best line that fits the training data. Let's make two predictions using the model's `predict()` function.

```
>>> print(model.predict([ [127], [248] ]))
[[ 438.94308857, 127.14839521]]
```

The model returned an array of predictions, one prediction for each input array. The first input, `[127]` , got a prediction of `438.94308857` . The seconds input, `[248]` , got a prediction of `127.14839521` . The reason for predicting on an array like `[127]` and not just `127` , is because you can have a model that makes a prediction using multiple features. We'll go over using multiple variables in linear regression later in this lesson. For now, let's stick to a single value.

Linear Regression Quiz

In this quiz, you'll be working with data on the average life expectancy at birth and the average BMI for males across the world. The data comes from Gapminder.

The data file can be found under the "bmi_and_life_expectancy.csv" tab in the quiz below. It includes three columns, containing the following data:

- **BMI** – The mean BMI of males in that country.

You'll need to complete each of the following steps:

1. Load the data

- The data is in the file called "bmi_and_life_expectancy.csv".
- Use pandas `read_csv` to load the data into a dataframe (don't forget to import pandas!)
- Assign the dataframe to the variable `bmi_life_data`.

2. Build a linear regression model

- Create a regression model using scikit-learn's `LinearRegression` and assign it to `bmi_life_model`.
- Fit the model to the data.

3. Predict using the model

- Predict using a BMI of 21.07931 and assign it to the variable `laos_life_exp`.

gapminder1.py bmi_and_life_expectancy.csv solution.py

```
# TODO: Add import statements
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression

# Assign the dataframe to this variable.
# TODO: Load the data

bmi_life_data = pd.read_csv('bmi_and_life_expectancy.csv')
#print(bmi_life_data)

y = bmi_life_data['Life expectancy']

X = bmi_life_data.drop(['Life expectancy', 'Country'], axis =1 )

print(X)


# Make and fit the linear regression model
#TODO: Fit the model and Assign it to bmi_life_model

model = LinearRegression()

bmi_life_model = model.fit(X,y)
```

```
laos_life_exp = model.predict([21.07931])

```

```
/usr/local/lib/python2.7/dist-packages/sklearn/utils/validation.py:386:
DeprecationWarning: Passing 1d arrays as data is deprecated in 0.17 and
willraise ValueError in 0.19. Reshape your data either using
X.reshape(-1, 1) if your data has a single feature or X.reshape(1, -1)
if it contains a single sample.
  DeprecationWarning)
          BMI
0    20.62058
1    26.44657
2    24.59620
3    27.63048
4    22.25083
5    25.35542
6    27.56373
7    26.46741
8    25.65117
```

RESET QUIZ TEST RUN SUBMIT ANSWER

NEXT