Multiple Linear Regression

In the last section, you saw how we can predict life expectancy using BMI. Here, BMI was the **predictor**, also known as an independent variable. A predictor is a variable you're looking at in order to make predictions about other variables, while the values you are trying to predict are known as dependent variables. In this case, life expectancy was the dependent variable.

Now, let's say we get new data on each person's heart rate as well. Can we create a prediction of life expectancy using both BMI and heart rate?

Absolutely! As we saw in the previous video, we can do that using multiple linear regression.

If the outcome you want to predict depends on more than one variable, you can make a more complicated model that takes this into account. As long as they're relevant to the situation, using more independent/predictor variables can help you get a better prediction.

When there's just one predictor, the linear regression model is a line, but as you add more predictor variables, you're adding more dimensions to the picture.
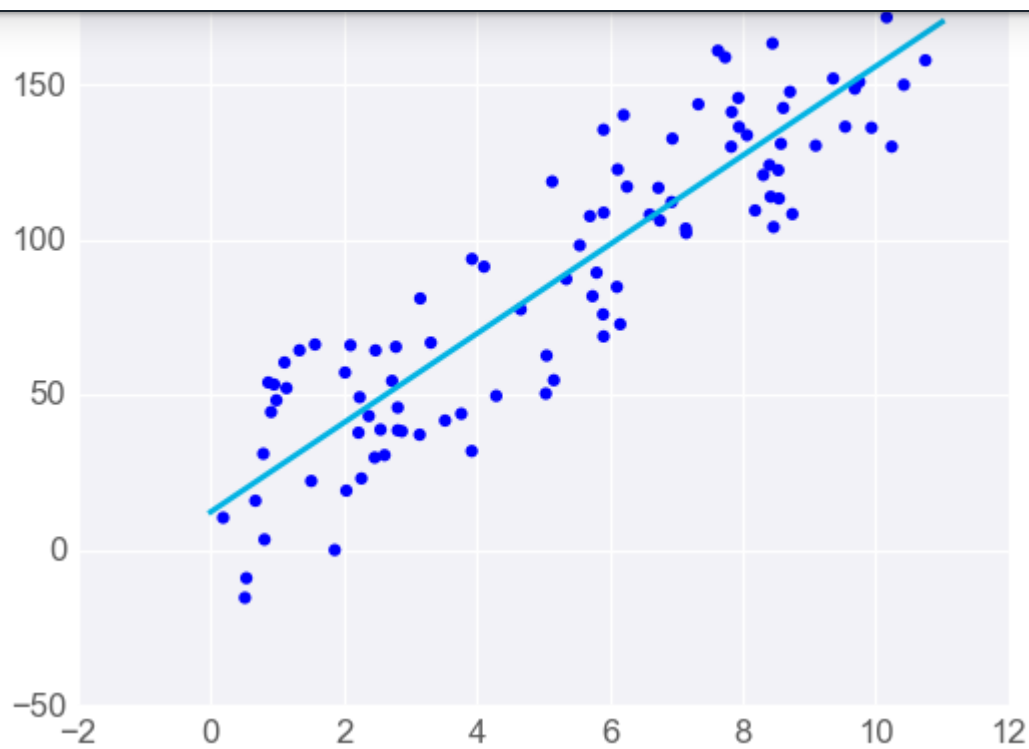
When you have one predictor variable, the equation of the line is

$$y = mx + b$$

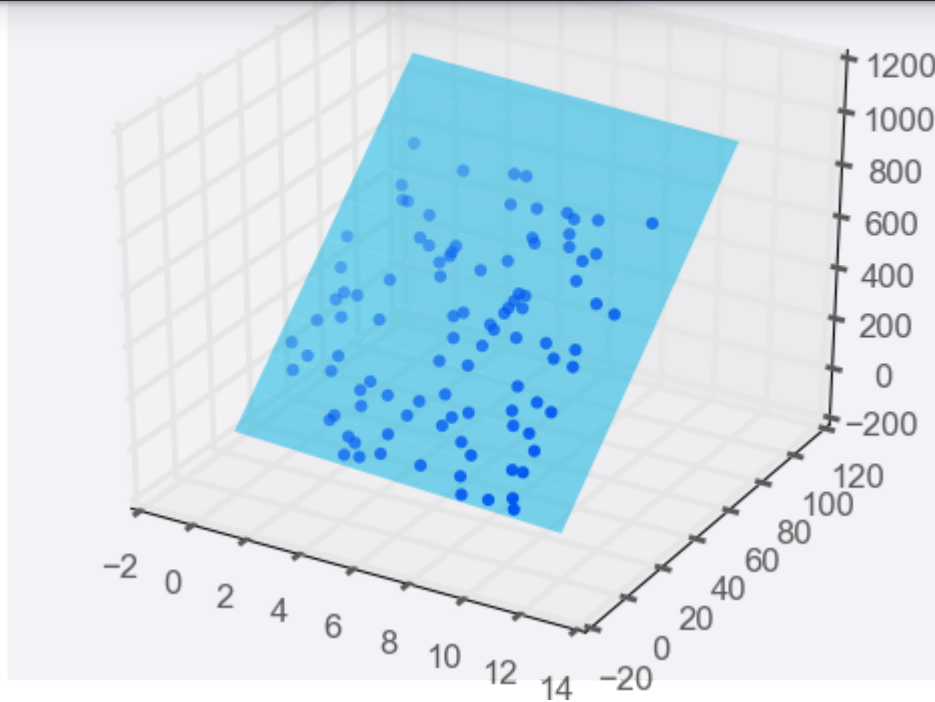and the plot might look something like this:

Linear regression with one predictor variable

Adding a predictor variable to go to two predictor variables means that the predicting equation is:

$$y = m_1 x_1 + m_2 x_2 + b$$

To represent this graphically, we'll need a three-dimensional plot, with the linear regression model represented as a plane:

Linear regression with two predictor variables

You can use more than two predictor variables, in fact you should use as many as is useful! If you use n predictor variables, then the model can be represented by the equation

$$y = m_1 x_1 + m_2 x_2 + m_3 x_3 + ... + m_n x_n + b$$

As you make a model with more predictor variables, it becomes harder to visualise, but luckily, everything else about linear regression stays the same. We can still fit models and make predictions in exactly the same way - time to try it!

Programming Quiz: Multiple Linear Regression

In this quiz, you'll be using the Boston house-prices dataset. The dataset consists of 13 features of 506 houses and their median value in $1000's. You'll fit a model on the 13 features to predict on the value of houses.

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

1. Build a linear regression model

- Fit the model to the data.

2. Predict using the model

- Predict the value of `sample_house` .

| quiz.py | solution.py |

```python
1   from sklearn.linear_model import LinearRegression
2   from sklearn.datasets import load_boston
3
4   # Load the data from the the boston house-prices dataset
5   boston_data = load_boston()
6   x = boston_data['data']
7   y = boston_data['target']
8
9   # Make and fit the linear regression model
10  # TODO: Fit the model and Assign it to the model variable
11  model = LinearRegression()
12
13  model.fit(x,y)
14
15  # Make a prediction using the model
16  sample_house = [[2.29690000e-01, 0.00000000e+00, 1.05900000e+01, 0.00000000e+00,
17                   6.32600000e+00, 5.25000000e+01, 4.35490000e+00, 4.00000000e+00,
18                   1.86000000e+01, 3.94870000e+02, 1.09700000e+01]]
19  # TODO: Predict housing price for the sample_house
20  prediction = model.predict(sample_house)
```

```
23.6842056923 is the correct prediction!
```

RESET QUIZ   TEST RUN   SUBMIT ANSWER

NEXT