

Lecture 3 part 3

July 16, 2021

```
[15]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import datasets
plt.rcParams['figure.figsize'] = [8,4]

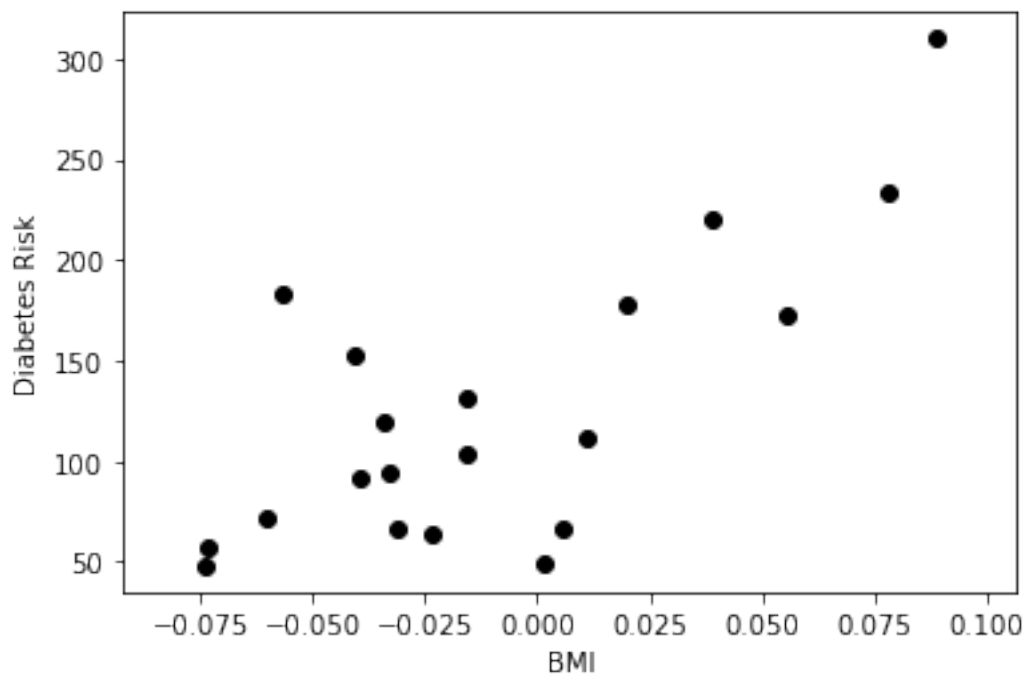
[2]: X,y = datasets.load_diabetes(return_X_y=True, as_frame = True)

[3]: X['one'] = 1

[4]: #Collect 20 data points
X_train = X.iloc[-20:]
y_train = y.iloc[-20:]

[5]: plt.scatter(X_train.bmi, y_train, color='black')
plt.xlabel('BMI')
plt.ylabel('Diabetes Risk')

[5]: Text(0, 0.5, 'Diabetes Risk')
```

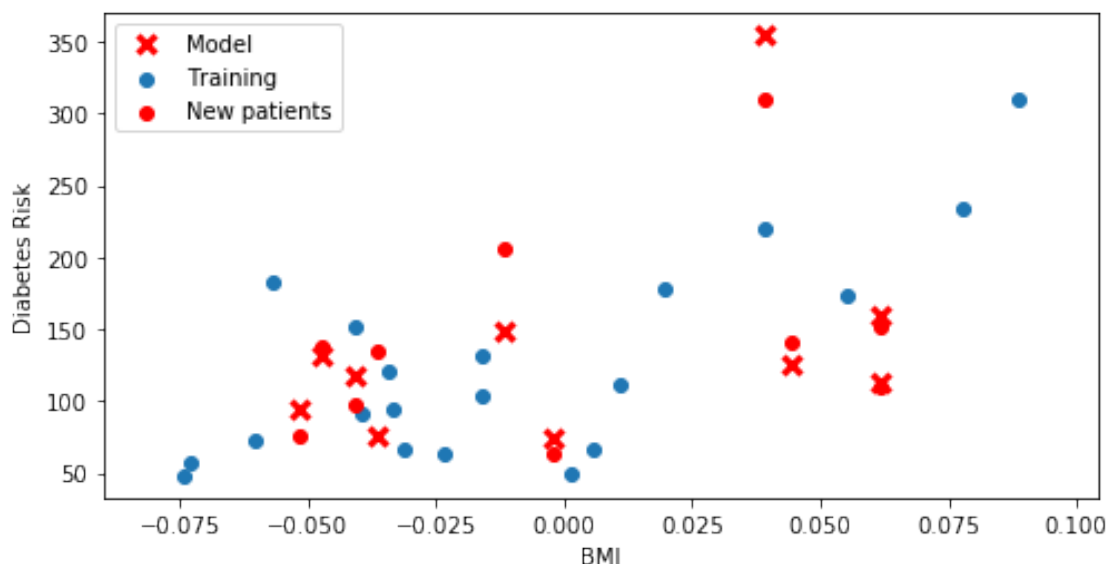


1 Ordinary Least Squares

```
[6]: #Derivative of MSE and equal to 0  
theta_best = np.linalg.inv(X_train.T.dot(X_train)).dot(X_train.T).dot(y_train)  
#Transforming it into a dataframe  
theta_best_df = pd.DataFrame(data=theta_best[np.newaxis,:], columns=X.columns)
```

2 Testing the model

```
[11]: X_test = X.iloc[:10]  
y_test = y.iloc[:10]  
  
y_test_pred = X_test.dot(theta_best)  
  
[27]: plt.scatter (X_train.bmi, y_train)  
plt.scatter (X_test.bmi, y_test, color='red', marker='o')  
plt.plot(X_test.bmi,y_test_pred, 'x', color='red', mew=3, markersize=8)  
plt.xlabel('BMI')  
plt.ylabel('Diabetes Risk')  
plt.legend(['Model', 'Training', 'New patients'])  
  
[27]: <matplotlib.legend.Legend at 0x7fd4f3236a90>
```



```
[ ]:
```