In [236…

```
import pandas as pd
import math
import numpy as np
from sklearn.model_selection import train_test_split, cross_val_score
import sklearn.preprocessing as pre
from sklearn.datasets import load_diabetes
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy_score, mean_squared_error
from sklearn.tree import DecisionTreeRegressor
```

In [237…

```
ds = load_diabetes()
X, y = ds.data, ds.target
print(X.shape)
print(y.shape)
```

```
(442, 10)
(442,)
```

In [238…

```
# Classifier has been set
dtrg = DecisionTreeRegressor()
```

In [239…

```
# Now we split data for training and testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, ran
```

In [240…

```
# Now we perform regression on it
dtrg_= dtrg.fit(X_train, y_train)
```

In [241…

```
print(cross_val_score(dtrg_, X, y=y))
```

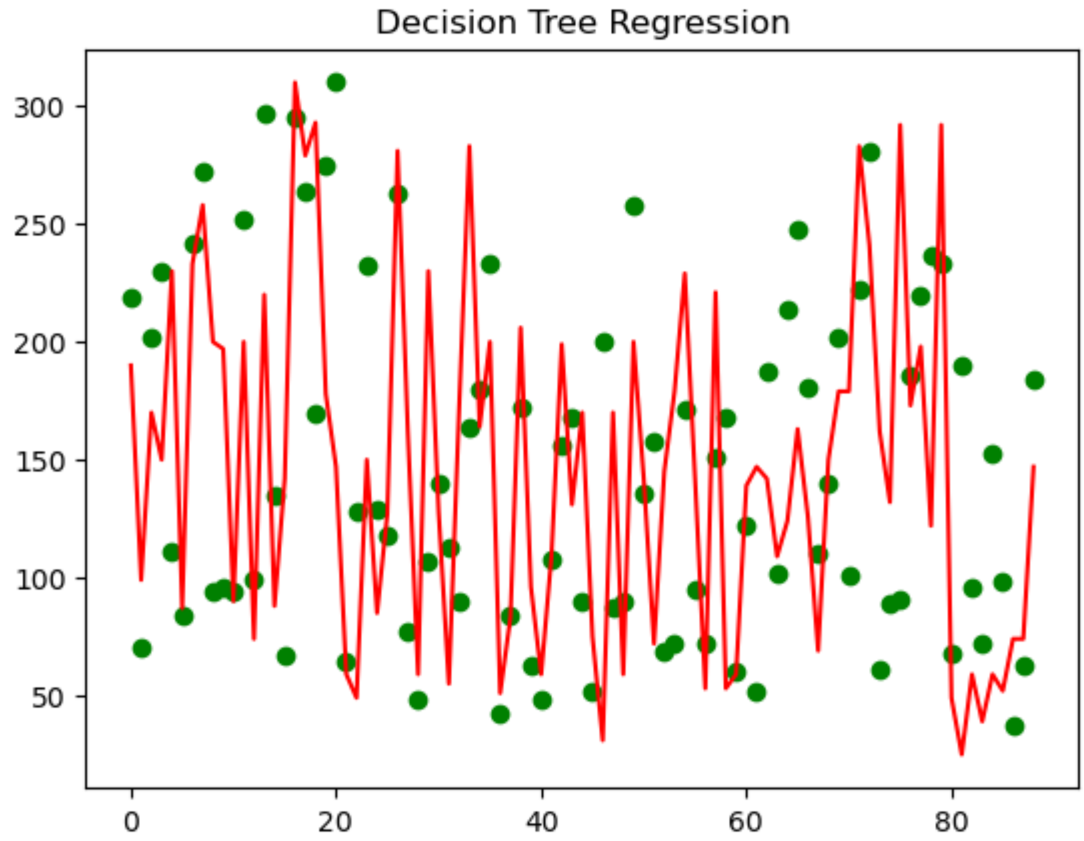
```
[-0.2622899  -0.03435629 -0.21331166  0.05334439 -0.10943631]
```

In [242…

```
hypothesis = dtrg_.predict(X_test)
```

In [243…

```
plt.figure()
plt.plot(y_test, 'og')
plt.plot(hypothesis, '-r')
plt.title('Decision Tree Regression')
plt.show()
```



In [244…

```
print(ds.DESCR)
```

```
.. _diabetes_dataset:

Diabetes dataset
----------------

Ten baseline variables, age, sex, body mass index, average blood
pressure, and six blood serum measurements were obtained for each of n =
442 diabetes patients, as well as the response of interest, a
quantitative measure of disease progression one year after baseline.

**Data Set Characteristics:**

  :Number of Instances: 442

  :Number of Attributes: First 10 columns are numeric predictive values

  :Target: Column 11 is a quantitative measure of disease progression one ye
ar after baseline

  :Attribute Information:
      - age     age in years
      - sex
      - bmi     body mass index
      - bp      average blood pressure
      - s1      tc, total serum cholesterol
      - s2      ldl, low-density lipoproteins
      - s3      hdl, high-density lipoproteins
      - s4      tch, total cholesterol / HDL
      - s5      ltg, possibly log of serum triglycerides level
      - s6      glu, blood sugar level

Note: Each of these 10 feature variables have been mean centered and scaled
by the standard deviation times the square root of `n_samples` (i.e. the sum
of squares of each column totals 1).

Source URL:
https://www4.stat.ncsu.edu/~boos/var.select/diabetes.html

For more information see:
Bradley Efron, Trevor Hastie, Iain Johnstone and Robert Tibshirani (2004) "L
east Angle Regression," Annals of Statistics (with discussion), 407-499.
(https://web.stanford.edu/~hastie/Papers/LARS/LeastAngle_2002.pdf)
```

In [ ]: