import matplotlib.pylab as plt  
import numpy as np  
%matplotlib inline  
from sklearn.linear\_model import LinearRegression  
from sklearn import datasets

* Khai báo thư viện

diabetes = datasets.load\_diabetes() # load data

* Load dữ liệu

diabetes.data.shape

(442, 10)

- Tạo hình ma trận

diabetes.target.shape

(442,)

- Ma trận vector

diabetes.feature\_names

['age', 'sex', 'bmi', 'bp', 's1', 's2', 's3', 's4', 's5', 's6']

- Các cột tên

from sklearn.model\_selection import train\_test\_split  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(diabetes.data, diabetes.target, test\_size=0.2, random\_state=0)

- Tạo dữ liệu thử

model = LinearRegression()

* Thiết lập model

model.fit(X\_train, y\_train)

LinearRegression(copy\_X=True, fit\_intercept=True, n\_jobs=None, normalize=False)

- Sử dụng fit

model.score(X\_test, y\_test)

0.33222203269065176

- Kiểm tra các điểm

model.coef\_

array([ -35.55683674, -243.1692265 , 562.75404632, 305.47203008,  
 -662.78772128, 324.27527477, 24.78193291, 170.33056502,  
 731.67810787, 43.02846824])

- Nhận các hệ số, beta

model.intercept\_

152.5381335195406

- Nhận giao điểm, c

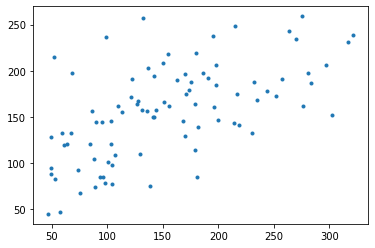
model.predict(X\_test)

array([238.47145247, 248.93170646, 164.05404165, 120.30794355,  
 187.42422054, 259.04865002, 113.55556372, 188.07597044,  
 149.49663441, 236.01099949, 172.24629506, 178.88073764,  
 109.15751983, 92.13508975, 243.33042043, 87.356971 ,  
 155.72606406, 66.99073989, 100.42610442, 218.09422877,  
 196.66287912, 161.29832968, 161.70779605, 156.52520454,  
 197.88796516, 167.57984206, 120.74478913, 84.83879727,  
 192.03728687, 160.60687024, 175.17178362, 84.22833237,  
 145.7995542 , 145.97333493, 140.96488953, 197.00421108,  
 165.94322494, 190.65906468, 128.22520508, 206.41941223,  
 84.35851196, 164.0256504 , 144.1056776 , 184.68355549,  
 177.80238966, 74.32855231, 143.3660286 , 138.67726085,  
 120.81146113, 234.34252077, 161.94390244, 74.5455476 ,  
 154.71905074, 156.78884927, 237.42227096, 174.23053048,  
 190.88212635, 118.98373473, 132.20418974, 168.52674824,  
 214.74245466, 171.42364091, 157.37409906, 108.86927343,  
 257.06329636, 152.17777143, 82.43686464, 231.56746032,  
 202.90641336, 47.18340199, 78.46954525, 129.30170908,  
 104.60253144, 144.65200281, 132.27974254, 190.04134164,  
 97.55541138, 197.51891007, 219.13709291, 186.13797012,  
 149.60913007, 208.42379455, 44.59036026, 206.20925368,  
 76.77377721, 94.94046865, 145.2955051 , 194.03776373,  
 132.78534336])

* Dự đoán dữ liệu không xác định

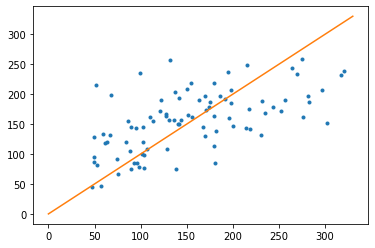
y\_pred = model.predict(X\_test)   
plt.plot(y\_test, y\_pred, '.')

[<matplotlib.lines.Line2D at 0x7fe854aaa048>]



* Dự đoán đồ thị và dữ liệu thực tế

y\_pred = model.predict(X\_test)   
plt.plot(y\_test, y\_pred, '.')  
x = np.linspace(0, 330, 100)  
y = x  
plt.plot(x, y)  
plt.show()



* Tạo đồ thị 1 đường, các dự đoán sẽ rơi trên đường đó