Practical - 15

AIM

Use some function for regularization of dataset based on problem 14.

PROBLEM

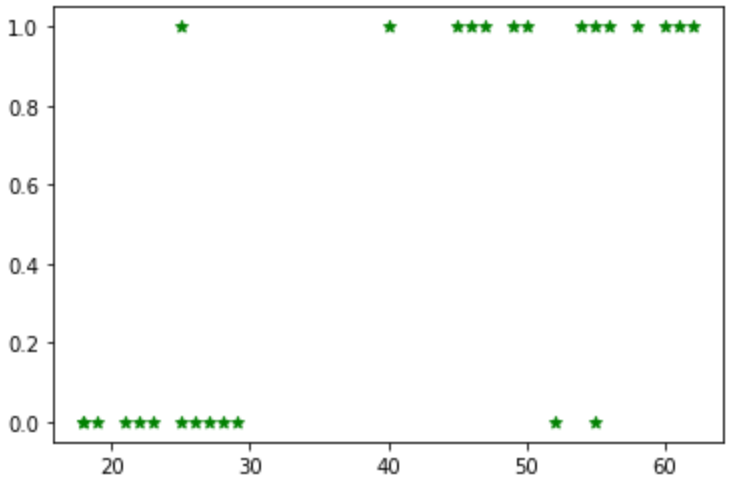
Based on age of a person predict whether a person will buy insurance or not.

CODE & OUTPUT

In [17]:

```
# import statements
import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline

# read the data file
df = pd.read_csv("insurance_data.csv")
df.head()

# scatter plot the data
plt.scatter(df.age, df.bought_insurance, marker="*", color="green")

# split the data into 2 parts: train data & test data
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(df[['age']], df.bought_insurance, trai

from sklearn.linear_model import RidgeClassifier
ridge_model = RidgeClassifier(alpha = 50)

# fit the training data on the model
ridge_model.fit(x_train, y_train)

# predict y for the test data
y_pred = ridge_model.predict(x_test)
print(y_pred)

# get the coeffient in the equation y = mx + c
print(ridge_model.coef_)

# get the intercept in the equation y = mx + c
print(ridge_model.intercept_)
```

```
[0 1 0 1 1 1]
[[0.04790596]]
[-1.99293978]
```



In [18]:

```
# the best fit line can be written as
y = 0.05 * df.age - 1.99
plt.plot(df.age, y, color="blue")
plt.scatter(df.age, df.bought_insurance, marker="*", color="green")

print(ridge_model.score(x_train, y_train))
print(ridge_model.score(x_test, y_test))
```

```
0.8571428571428571
0.8333333333333334
```