parth.pandey13103347@gmail.com_10_C

June 21, 2020

0.1 Task-C: Regression outlier effect.

Objective:Visualization best fit linear regression line for different scenarios

0.1.1 Imports

[1]:
```
# you should not import any other packages
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings("ignore")
import numpy as np
from sklearn.linear_model import SGDRegressor
from IPython.display import Latex
```

0.1.2 Data Creation Functions

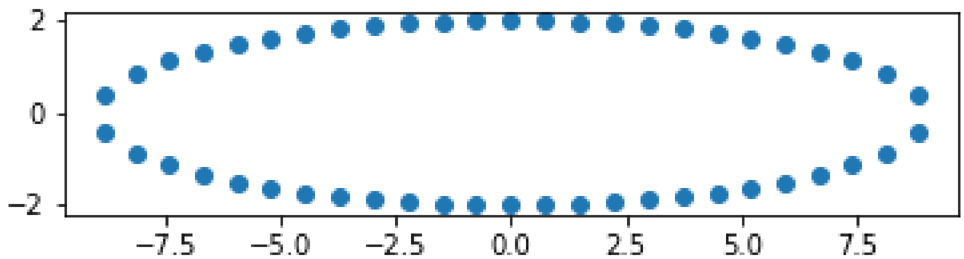
[176]:
```
import numpy as np
import scipy as sp
import scipy.optimize

def angles_in_ellipse(num,a,b):
    assert(num > 0)
    assert(a < b)
    angles = 2 * np.pi * np.arange(num) / num
    if a != b:
        e = (1.0 - a ** 2.0 / b ** 2.0) ** 0.5
        tot_size = sp.special.ellipeinc(2.0 * np.pi, e)
        arc_size = tot_size / num
        arcs = np.arange(num) * arc_size
        res = sp.optimize.root(
            lambda x: (sp.special.ellipeinc(x, e) - arcs), angles)
        angles = res.x
    return angles
```

[177]:
```
a = 2
b = 9
n = 50
```

```
phi = angles_in_ellipse(n, a, b)
e = (1.0 - a ** 2.0 / b ** 2.0) ** 0.5
arcs = sp.special.ellipeinc(phi, e)

fig = plt.figure()
ax = fig.gca()
ax.axes.set_aspect('equal')
ax.scatter(b * np.sin(phi), a * np.cos(phi))
plt.show()
```



0.1.3 Generating Data

[178]:
```
x_train= b * np.sin(phi)
y_train= a * np.cos(phi)
```

0.1.4 Steps to achieve Objective

0.1.5 Observation on checking the documentation of loss functions of Linear Models.

- On checking out the https://scikit-learn.org/stable/modules/linear_model.html

  It was brought into light that the regression implementation with regularization is called **ridge regression**

- **"This classifier is sometimes referred to as a Least Squares Support Vector Machines with a linear kernel."**

  Here the documentation says that Linear SVM is pretty similar to the current implementation.

0.1.6 Formula Used in the Self Implementation of Linear Regression

Linear Regression Loss Function Used

- Loss Function

  $\min_w (1/N) * \|X.w - y\|_2^2 + \alpha\|w\|_2^2$

- Gradient for w

  $\delta L/\delta w_{w_{old}} = 2 * X * ((w.X) + b - y) + 2\alpha w$

- Gradient for b

  $\delta L/\delta b_{b_{old}} = 2 * ((w.X) + b - y)$

Defining Functions for the above formula

[179]:
```
# Weight Intialization
def init_weights(x_train):
    w = np.zeros_like(x_train)
    b = 0
    return w,b

def sigmoid(z):
    return 1/(1+ np.exp(-z))

# gradient function
# Loss = Square Loss
def gradient_dw(x, y, w, b, param, N):
    return  x * ((np.dot(w,x) + b) - y) * (1/N) + 2 * (param) * w

def gradient_db(x, y, w, b, param, N):
    return ((np.dot(w,x)  + b) - y)/N

def loss(y_true,y_pred):
    return np.sum((y_true- y_pred)**2)

def draw_line(coef,intercept, mi, ma, ax, color='green', label=''):
    # for the separating hyper plane ax+by+c=0, the weights are [a, b] and the␣
↪intercept is c
    # to draw the hyper plane we are creating two points
    # 1. ((b*min-c)/a, min) i.e ax+by+c=0 ==> ax = (-by-c) ==> x = (-by-c)/a␣
↪here in place of y we are keeping the minimum value of y
    # 2. ((b*max-c)/a, max) i.e ax+by+c=0 ==> ax = (-by-c) ==> x = (-by-c)/a␣
↪here in place of y we are keeping the maximum value of y
    points=np.array([(mi, (coef*mi + intercept)),(ma, (coef*ma + intercept))])
    ax.plot(points[:,0], points[:,1], color=color, label=label)
```

Defining train function

[180]:
```
def train(x_train, y_train, epochs, param, eta):
    w,b = init_weights(x_train[0])
    epoch_loss = 0
    N = x_train.shape[0]
    for epoch in range(1,epochs+1):
        for x, y in zip(x_train, y_train):
            dw = gradient_dw(x,y,w,b,param, N)
            db = gradient_db(x,y,w,b,param, N)
```

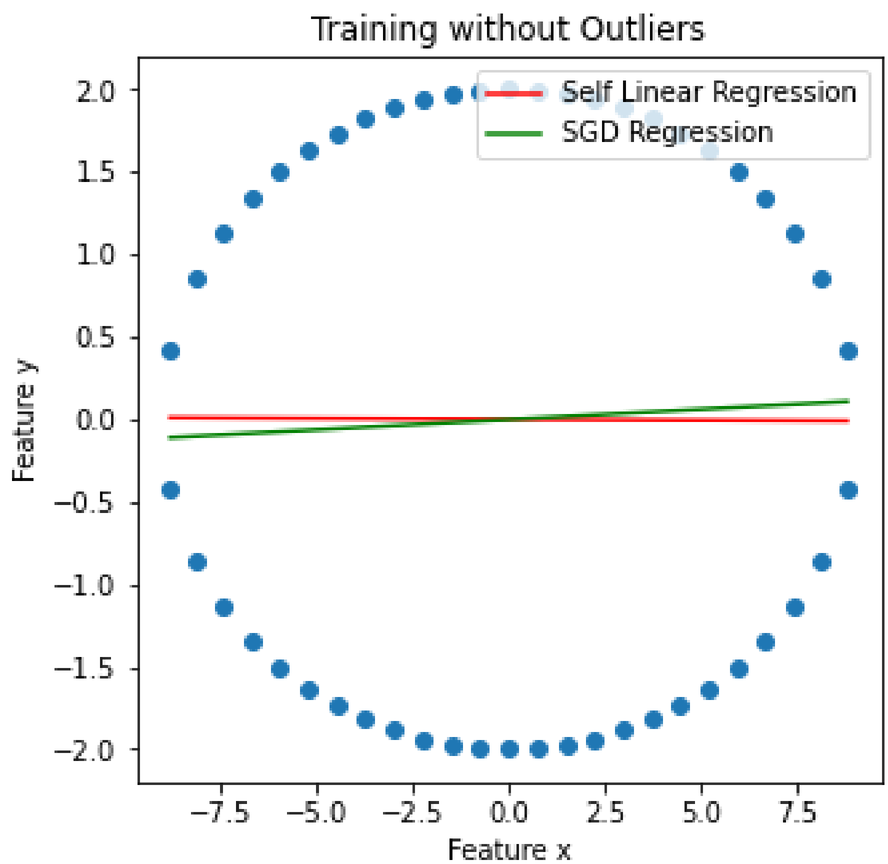
```
            w = w - (eta * dw)
            b = b - (eta * db)
        y_pred = []
        for x in x_train:
            y_pred.append(np.dot(w,x) + b)
        old_loss = epoch_loss
        epoch_loss = loss(y_train, y_pred)
        if epoch_loss==old_loss:
            break
    return w,b
```

0.1.7 Linear Regression Implementation

Training with Original Data

[181]:

```
# training on original data
_, ax = plt.subplots(1,1,figsize=(5,5))
w, b = train(x_train, y_train, 1000, 0.0001, 0.001)
clf =SGDRegressor(loss='squared_loss', alpha=0.0001, eta0=0.001,␣
 ↪learning_rate='constant', random_state=0)
clf.fit(x_train.reshape(-1,1), y_train)
ax.scatter(x_train, y_train)
draw_line(w,b,np.min(x_train),np.max(x_train),ax,color='red', label='Self␣
 ↪Linear Regression')
draw_line(clf.coef_[0],clf.intercept_[0],np.min(x_train),np.max(x_train),ax,␣
 ↪label='SGD Regression')
ax.set_title('Training without Outliers')
ax.legend(loc=1)
ax.set_xlabel('Feature x')
ax.set_ylabel('Feature y')
plt.show()
```

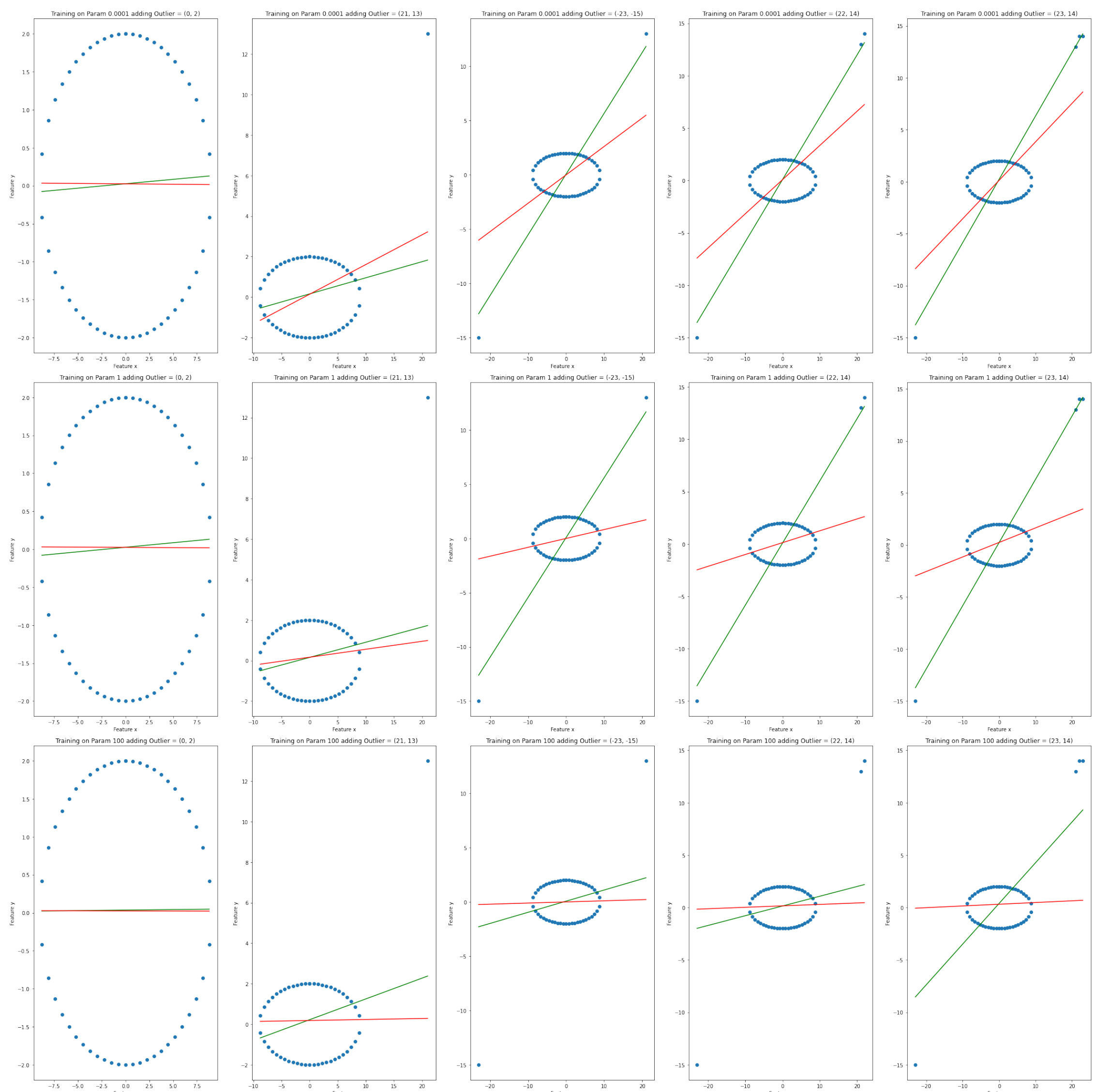


Training with outliers

[182]:
```
outliers = [(0,2),(21, 13), (-23, -15), (22,14), (23, 14)]
hyperparams = [0.0001, 1, 100]
```

[183]:
```
_, ax = plt.subplots(3,5,figsize=(30,30))
for i, param in enumerate(hyperparams):
    x_train_temp = x_train
    y_train_temp = y_train
    for j, ele in enumerate(outliers):
        x_train_temp = np.append(x_train_temp, ele[0])
        y_train_temp = np.append(y_train_temp, ele[1])
        w,b = train(x_train_temp, y_train_temp, 1000, param, 0.001)
        clf = SGDRegressor(loss='squared_loss', alpha=param, random_state=0)
        clf.fit(x_train_temp.reshape(-1,1), y_train_temp)
        ax[i][j].scatter(x_train_temp, y_train_temp)
        draw_line(clf.coef_[0],clf.intercept_[0],np.min(x_train_temp),np.
 ↪max(x_train_temp),ax[i][j], label='SGD Regression')
```

```
        draw_line(w,b,np.min(x_train_temp),np.max(x_train_temp),ax[i][j],␣
↪color='red', label='Self Linear Regression')
        ax[i][j].set_title('Training on Param {} adding Outlier = {}'.
↪format(param, ele))
        ax[i][j].set_xlabel('Feature x')
        ax[i][j].set_ylabel('Feature y')
plt.tight_layout()
plt.show()
```



0.1.8 Observations based on self implementation of SGDRegressor

- As the hyperparameter is increasing the hyperplane becomes less sensitive towards the outliers but the model still shifts towards the outliers but only by a small margin.\

- Reason for happening is \

$$w_n = w_o - \frac{w_o^t x + b}{N} + \frac{y}{N} - 2\lambda w_o$$

as the value of λ increases there is more stringent increase in the value of w_n , hence weights are much less affected in presence of an outlier.\