Worksheet 21

Name: Hao Qi, Hui Zheng
UID: U96305250, U80896784

Topics

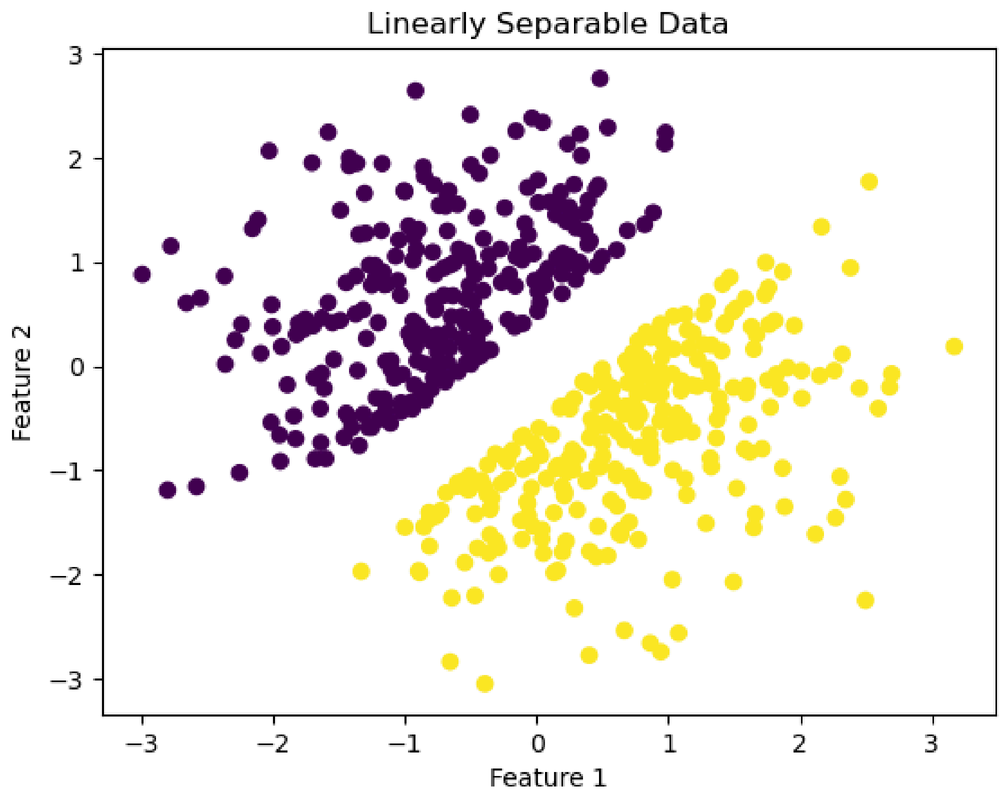
- Logistic Regression
- Gradient Descent

Logistic Regression

In [ ]:
```
import numpy as np
import matplotlib.pyplot as plt
import sklearn.datasets as datasets
from sklearn.pipeline import make_pipeline
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import PolynomialFeatures

centers = [[0, 0]]
t, _ = datasets.make_blobs(n_samples=750, centers=centers, cluster_std=1, random_state=(

# LINE
def generate_line_data():
    # create some space between the classes
    X = np.array(list(filter(lambda x : x[0] - x[1] < -.5 or x[0] - x[1] > .5, t)))
    Y = np.array([1 if x[0] - x[1] >= 0 else 0 for x in X])
    return X, Y

# CIRCLE
def generate_circle_data(t):
    # create some space between the classes
    X = np.array(list(filter(lambda x : (x[0] - centers[0][0])**2 + (x[1] - centers[0][(
    Y = np.array([1 if (x[0] - centers[0][0])**2 + (x[1] - centers[0][1])**2 >= 1 else (
    return X, Y

# XOR
def generate_xor_data():
    X = np.array([
        [0,0],
        [0,1],
        [1,0],
        [1,1]])
    Y = np.array([x[0]^x[1] for x in X])
    return X, Y
```

a) Using the above code, generate and plot data that is linearly separable.

In [ ]:
```
X, Y = generate_line_data()

plt.scatter(X[:, 0], X[:, 1], c=Y)
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.title('Linearly Separable Data')
plt.show()
```



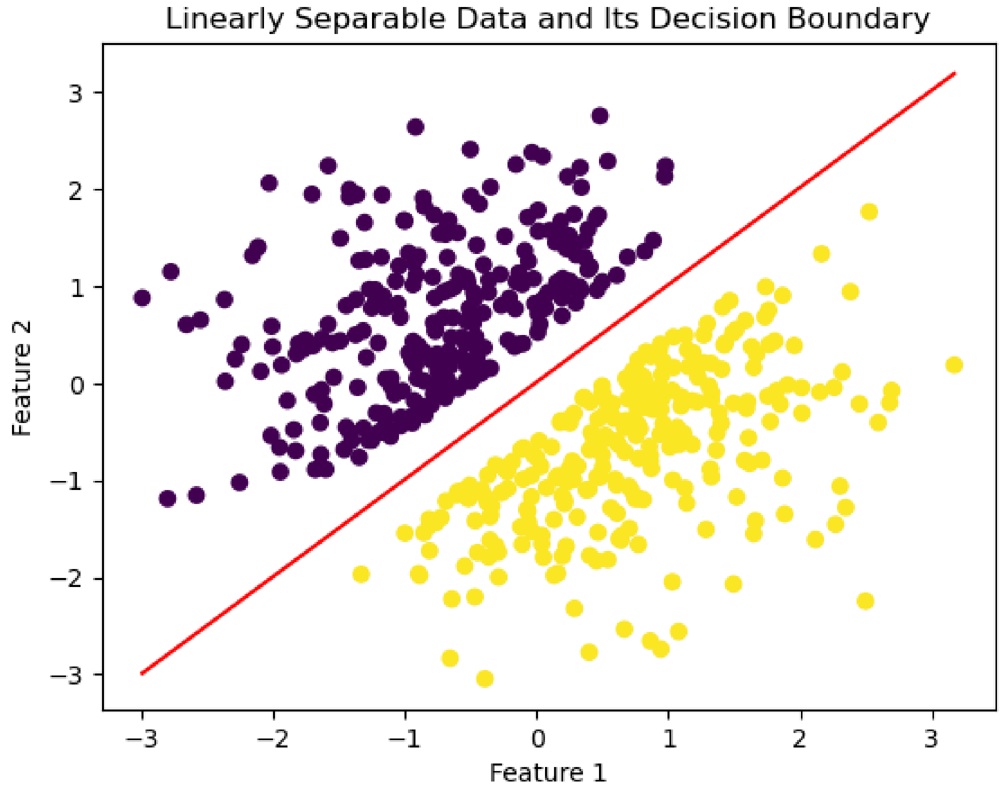
b) Fit a logistic regression model to the data and print out the coefficients.

```
In [ ]: model = LogisticRegression().fit(X, Y)
        print(f"coefficients: {model.coef_[0]}")
        print(f"intercept: {model.intercept_}")
```

```
coefficients: [ 4.11337993 -4.10105513]
intercept: [0.05839469]
```

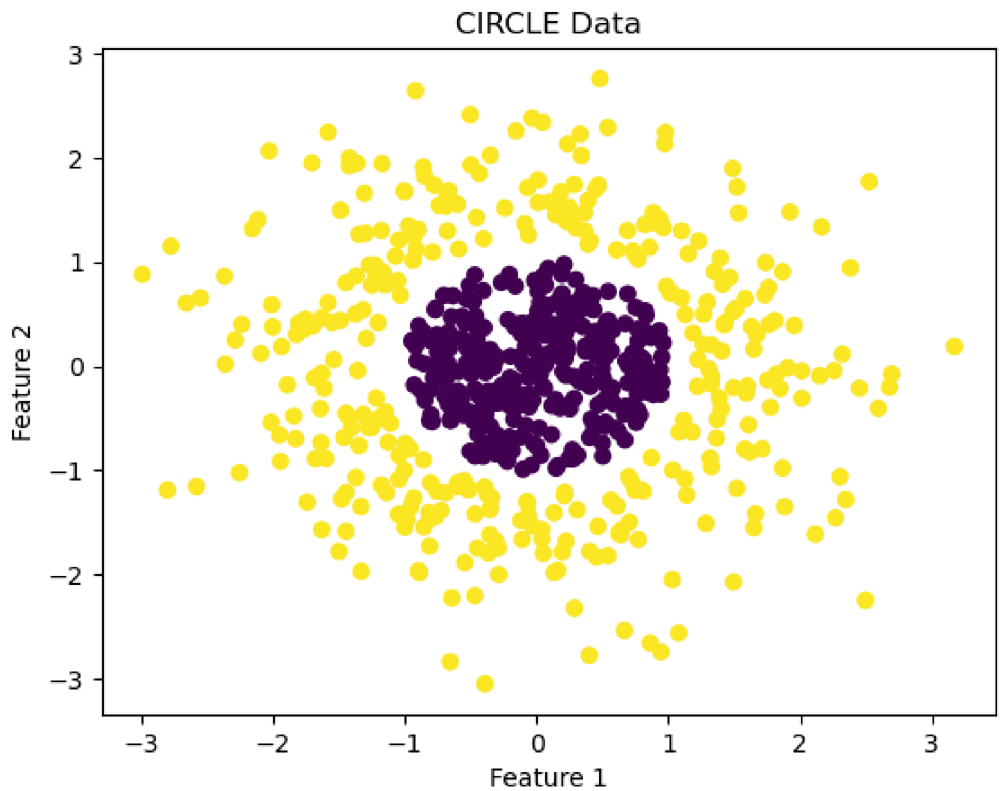
c) Using the coefficients, plot the line through the scatter plot you created in a). (Note: you need to do some math to get the line in the right form)

```
In [ ]: # c1 * x1 + c2 * x2 + intercept = 0
        x_values = np.linspace(np.min(X[:, 0]), np.max(X[:, 0]), 10)
        y_values = -(model.coef_[0][0]/model.coef_[0][1]) * x_values - model.intercept_[0]/mode

        plt.scatter(X[:, 0], X[:, 1], c=Y)
        plt.plot(x_values, y_values, c='r')
        plt.xlabel('Feature 1')
        plt.ylabel('Feature 2')
        plt.title('Linearly Separable Data and Its Decision Boundary')
        plt.show()
```



d) Using the above code, generate and plot the CIRCLE data.

In [ ]:
```
X, Y = generate_circle_data(t)

plt.scatter(X[:, 0], X[:, 1], c=Y)
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.title('CIRCLE Data')
plt.show()
```



e) Notice that the equation of an ellipse is of the form

$$ax^2 + by^2 = c$$

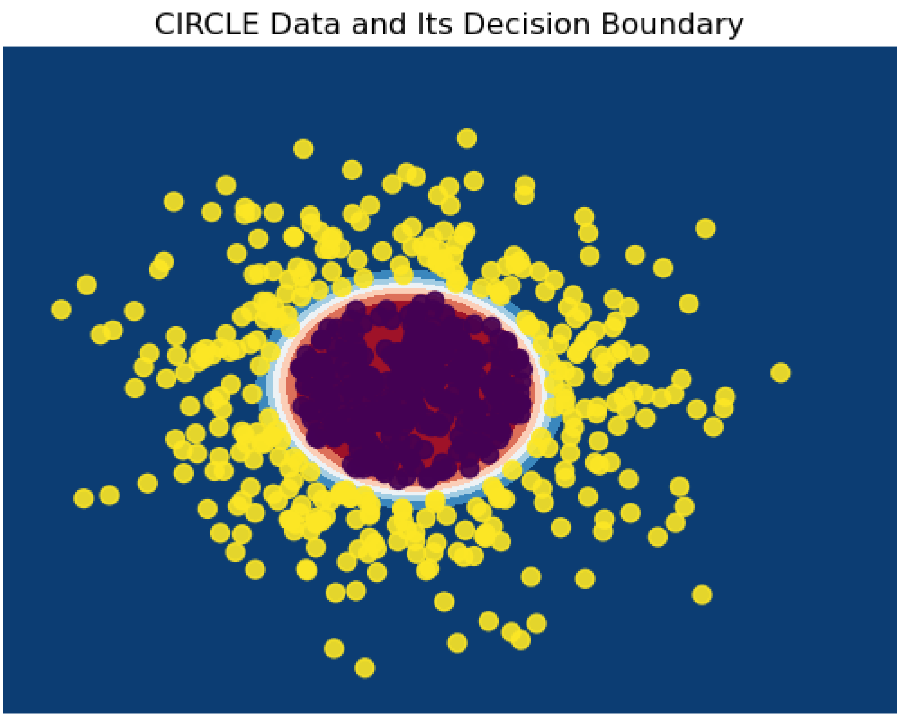
Fit a logistic regression model to an appropriate transformation of X.

In [ ]:
```
poly = PolynomialFeatures(degree=2, include_bias=False)
lr = LogisticRegression()
model = make_pipeline(poly, lr).fit(X, Y)
print(f"coefficients: {lr.coef_[0]}")   # x, y, x^2, xy, y^2
print(f"intercept: {lr.intercept_}")
```

```
coefficients: [ 0.02985162 -0.04753247  4.90954898  0.37928     4.95645605]
intercept: [-6.47659385]
```

f) Plot the decision boundary using the code below.

In [ ]:
```
# create a mesh to plot in
h = .02  # step size in the mesh
x_min, x_max = X[:, 0].min() - .5, X[:, 0].max() + 1
y_min, y_max = X[:, 1].min() - .5, X[:, 1].max() + 1
xx, yy = np.meshgrid(np.arange(x_min, x_max, h),
                     np.arange(y_min, y_max, h))
meshData = np.c_[xx.ravel(), yy.ravel()]

fig, ax = plt.subplots()
A = model.predict_proba(meshData)[:, 1].reshape(xx.shape)
Z = model.predict(meshData).reshape(xx.shape)
ax.contourf(xx, yy, A, cmap="RdBu", vmin=0, vmax=1)
ax.axis('off')

# plot also the training points
ax.scatter(X[:, 0], X[:, 1], c=Y, s=50, alpha=0.9)
plt.title('CIRCLE Data and Its Decision Boundary')
plt.show()
```

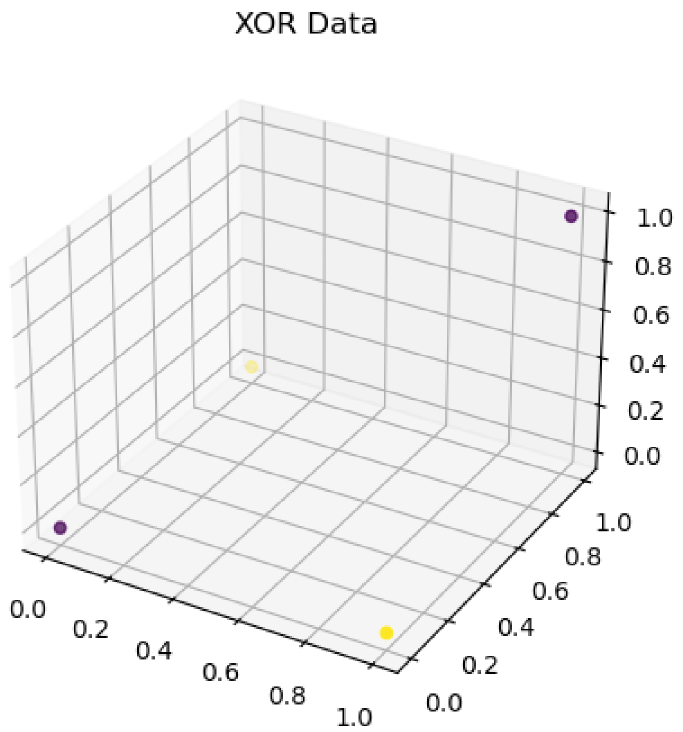


g) Plot the XOR data. In this 2D space, the data is not linearly separable, but by introducing a new feature

$$x_3 = x_1 * x_2$$

(called an interaction term) we should be able to find a hyperplane that separates the data in 3D. Plot this new dataset in 3D.

In [ ]:
```
from mpl_toolkits.mplot3d import Axes3D

X, Y = generate_xor_data()
ax = plt.axes(projection='3d')
ax.scatter3D(X[:, 0], X[:, 1], X[:, 0]*X[:, 1], c=Y)
plt.title('XOR Data')
plt.show()
```



h) Apply a logistic regression model using the interaction term. Plot the decision boundary.

In [ ]:
```
poly = PolynomialFeatures(interaction_only=True)
lr = LogisticRegression(verbose=0)
model = make_pipeline(poly, lr).fit(X, Y)

# create a mesh to plot in
h = .02  # step size in the mesh
x_min, x_max = X[:, 0].min() - .5, X[:, 0].max() + 1
y_min, y_max = X[:, 1].min() - .5, X[:, 1].max() + 1
xx, yy = np.meshgrid(np.arange(x_min, x_max, h),
                     np.arange(y_min, y_max, h))
meshData = np.c_[xx.ravel(), yy.ravel()]

fig, ax = plt.subplots()
A = model.predict_proba(meshData)[:, 1].reshape(xx.shape)
Z = model.predict(meshData).reshape(xx.shape)
ax.contourf(xx, yy, A, cmap="RdBu", vmin=0, vmax=1)
ax.axis('off')

# plot also the training points
ax.scatter(X[:, 0], X[:, 1], color=Y, s=50, alpha=0.9)
plt.title('XOR Data and Its Decision Boundary')
plt.show()
```

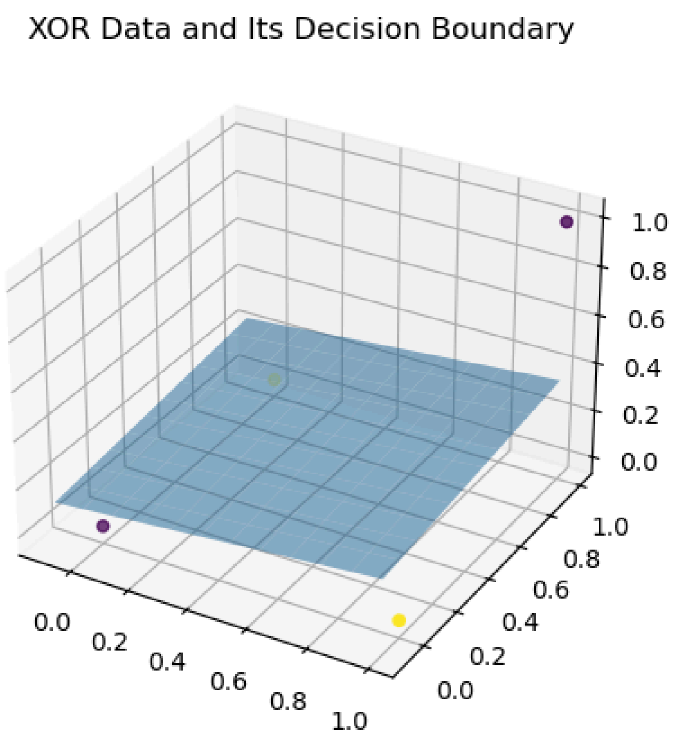
XOR Data and Its Decision Boundary

In [ ]:

```
%matplotlib widget
for i in range(20000):
    for solver in ['lbfgs', 'liblinear', 'newton-cg', 'newton-cholesky', 'sag', 'saga']
        X_transform = PolynomialFeatures(interaction_only=True, include_bias=False).fit_
        model = LogisticRegression(verbose=0, solver=solver, random_state=i, max_iter=1(
        model.fit(X_transform, Y)
        score = model.score(X_transform, Y)
        if score > .75:
            print("random state = ", i)
            print("solver = ", solver)
            print("score = ", score)
            break

print(model.coef_)
print(model.intercept_)

xx, yy = np.meshgrid([x / 10 for x in range(-1, 11)], [x / 10 for x in range(-1, 11)])
z = - model.intercept_ / model.coef_[0][2] - model.coef_[0][0] * xx / model.coef_[0][2]

ax = plt.axes(projection='3d')
ax.scatter3D(X[:, 0], X[:, 1], X[:, 0]*X[:, 1], c=Y)
ax.plot_surface(xx, yy, z, alpha=0.5)
plt.title('XOR Data and Its Decision Boundary')
plt.show()
```

```
[[ 0.04307198  0.04306862 -0.43033115]]
[0.06381617]
```

Figure



i) Using the code below that generates 3 concentric circles, fit a logisitc regression model to it and plot the decision boundary.

In [ ]:

```
t, _ = datasets.make_blobs(n_samples=1500, centers=centers, cluster_std=2,
                                random_state=0)

# CIRCLES
def generate_circles_data(t):
    def label(x):
        if x[0]**2 + x[1]**2 >= 2 and x[0]**2 + x[1]**2 < 8:
            return 1
        if x[0]**2 + x[1]**2 >= 8:
            return 2
        return 0
    # create some space between the classes
    X = np.array(list(filter(lambda x : (x[0]**2 + x[1]**2 < 1.8 or x[0]**2 + x[1]**2 >
    Y = np.array([label(x) for x in X])
    return X, Y

X, Y = generate_circles_data(t)

poly = PolynomialFeatures(2)
lr = LogisticRegression(max_iter=500, verbose=2)
model = make_pipeline(poly, lr).fit(X, Y)

# create a mesh to plot in
h = .02
x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1
y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1
xx, yy = np.meshgrid(np.arange(x_min, x_max, h), np.arange(y_min, y_max, h))
meshData = np.c_[xx.ravel(), yy.ravel()]

Z = model.predict(meshData).reshape(xx.shape)

fig, ax = plt.subplots()
```

```python
ax.contourf(xx, yy, Z, cmap='RdBu', vmin=0, vmax=1, alpha=0.5)
ax.axis('off')
# plot also the training points
scatter = ax.scatter(X[:, 0], X[:, 1], c=Y, s=20, alpha=0.9)
plt.title('Concentric Data and Its Decision Boundary')
plt.show()
```

```
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
 This problem is unconstrained.
[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:    0.0s remaining:    0.0s
[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:    0.0s finished
```

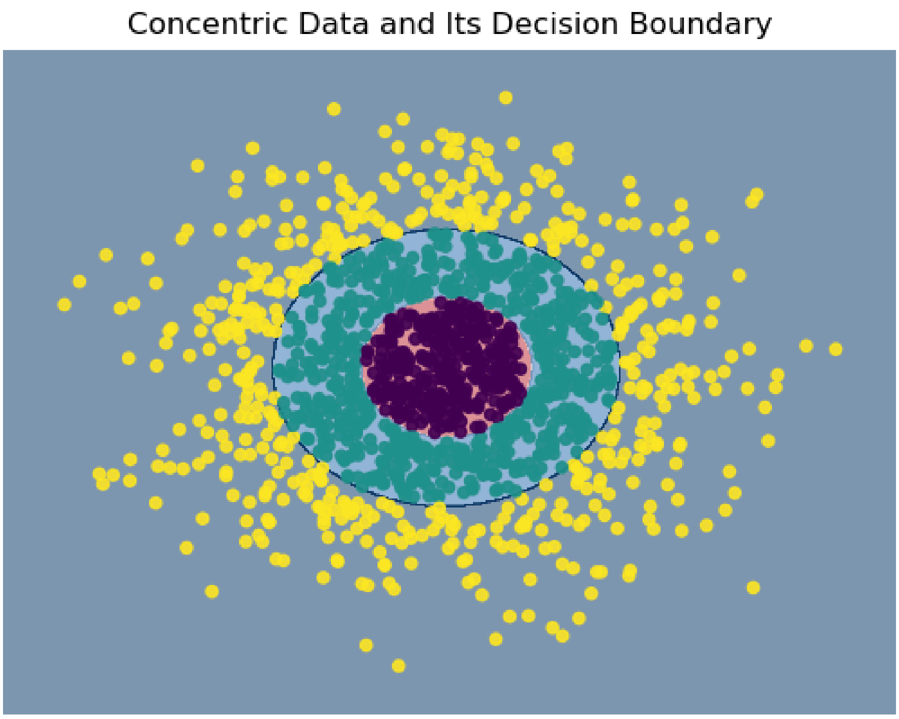
```
RUNNING THE L-BFGS-B CODE

           * * *

Machine precision = 2.220D-16
 N =           21     M =           10

At X0         0 variables are exactly at the bounds

At iterate    0    f=  1.54575D+03    |proj g|=  2.20592D+03

At iterate    1    f=  1.25980D+03    |proj g|=  7.85645D+02

At iterate    2    f=  1.16386D+03    |proj g|=  4.67390D+02

At iterate    3    f=  1.12014D+03    |proj g|=  8.53618D+02

At iterate    4    f=  1.07857D+03    |proj g|=  2.28644D+02

At iterate    5    f=  1.03868D+03    |proj g|=  2.46093D+02

At iterate    6    f=  8.75292D+02    |proj g|=  2.23884D+02

At iterate    7    f=  6.72984D+02    |proj g|=  1.69393D+02

At iterate    8    f=  4.06106D+02    |proj g|=  3.22571D+02

At iterate    9    f=  3.84632D+02    |proj g|=  4.67515D+02

At iterate   10    f=  2.26564D+02    |proj g|=  2.61932D+02

At iterate   11    f=  1.84398D+02    |proj g|=  1.15386D+02

At iterate   12    f=  1.54569D+02    |proj g|=  6.06745D+01

At iterate   13    f=  1.34331D+02    |proj g|=  4.43007D+01

At iterate   14    f=  1.24113D+02    |proj g|=  4.50198D+01

At iterate   15    f=  1.21743D+02    |proj g|=  3.63598D+01

At iterate   16    f=  1.19874D+02    |proj g|=  1.08601D+01

At iterate   17    f=  1.19500D+02    |proj g|=  6.08812D+00

At iterate   18    f=  1.19298D+02    |proj g|=  3.58407D+00

At iterate   19    f=  1.19015D+02    |proj g|=  6.65077D+00

At iterate   20    f=  1.18680D+02    |proj g|=  1.37452D+01

At iterate   21    f=  1.18079D+02    |proj g|=  1.98560D+01

At iterate   22    f=  1.16887D+02    |proj g|=  1.87388D+01

At iterate   23    f=  1.16408D+02    |proj g|=  2.74606D+01

At iterate   24    f=  1.14845D+02    |proj g|=  1.68476D+01

At iterate   25    f=  1.12048D+02    |proj g|=  1.01173D+01

At iterate   26    f=  1.09277D+02    |proj g|=  2.60636D+01

At iterate   27    f=  1.05058D+02    |proj g|=  4.27241D+01

At iterate   28    f=  9.92827D+01    |proj g|=  7.46190D+01
```

```
At iterate   29    f=  9.20571D+01    |proj g|=  5.06010D+01

At iterate   30    f=  8.70119D+01    |proj g|=  3.20989D+01

At iterate   31    f=  8.24104D+01    |proj g|=  8.74098D+00

At iterate   32    f=  8.08462D+01    |proj g|=  7.93679D+00

At iterate   33    f=  7.86450D+01    |proj g|=  1.85045D+01

At iterate   34    f=  7.57509D+01    |proj g|=  2.16495D+01

At iterate   35    f=  7.48794D+01    |proj g|=  2.70982D+01

At iterate   36    f=  7.30267D+01    |proj g|=  4.78293D+00

At iterate   37    f=  7.28149D+01    |proj g|=  2.79623D+00

At iterate   38    f=  7.26223D+01    |proj g|=  2.74530D+00

At iterate   39    f=  7.24812D+01    |proj g|=  4.17543D+00

At iterate   40    f=  7.24079D+01    |proj g|=  3.65780D+00

At iterate   41    f=  7.23176D+01    |proj g|=  2.28953D+00

At iterate   42    f=  7.22311D+01    |proj g|=  1.40639D+00

At iterate   43    f=  7.21318D+01    |proj g|=  1.93637D+00

At iterate   44    f=  7.20051D+01    |proj g|=  2.26014D+00

At iterate   45    f=  7.18225D+01    |proj g|=  1.89147D+00

At iterate   46    f=  7.17636D+01    |proj g|=  6.97182D+00

At iterate   47    f=  7.15057D+01    |proj g|=  6.28035D+00

At iterate   48    f=  7.13583D+01    |proj g|=  1.52391D+00

At iterate   49    f=  7.13213D+01    |proj g|=  1.17885D+00

At iterate   50    f=  7.12906D+01    |proj g|=  6.37428D-01

At iterate   51    f=  7.12722D+01    |proj g|=  1.65809D+00

At iterate   52    f=  7.12515D+01    |proj g|=  1.06709D+00

At iterate   53    f=  7.12293D+01    |proj g|=  1.29824D+00

At iterate   54    f=  7.10856D+01    |proj g|=  2.41904D+00

At iterate   55    f=  7.08821D+01    |proj g|=  2.78025D+00

At iterate   56    f=  7.07502D+01    |proj g|=  2.89623D+00

At iterate   57    f=  7.05884D+01    |proj g|=  1.27468D+00

At iterate   58    f=  7.05234D+01    |proj g|=  1.08223D+00

At iterate   59    f=  7.05117D+01    |proj g|=  1.17135D+00

At iterate   60    f=  7.05030D+01    |proj g|=  2.59818D-01

At iterate   61    f=  7.05006D+01    |proj g|=  2.21044D-01

At iterate   62    f=  7.04990D+01    |proj g|=  6.55160D-01
```

```
At iterate   63    f=  7.04963D+01    |proj g|=  3.54271D-01

At iterate   64    f=  7.04945D+01    |proj g|=  1.75421D-01

At iterate   65    f=  7.04930D+01    |proj g|=  1.71078D-01

At iterate   66    f=  7.04919D+01    |proj g|=  1.35809D-01

At iterate   67    f=  7.04904D+01    |proj g|=  1.66074D-01

At iterate   68    f=  7.04872D+01    |proj g|=  6.49632D-01

At iterate   69    f=  7.04820D+01    |proj g|=  4.94024D-01

At iterate   70    f=  7.04689D+01    |proj g|=  4.86061D-01

At iterate   71    f=  7.04669D+01    |proj g|=  3.11825D-01

At iterate   72    f=  7.04636D+01    |proj g|=  1.57499D-01

At iterate   73    f=  7.04623D+01    |proj g|=  2.12111D-01

At iterate   74    f=  7.04619D+01    |proj g|=  1.72148D-01

At iterate   75    f=  7.04616D+01    |proj g|=  6.92299D-02

At iterate   76    f=  7.04614D+01    |proj g|=  4.63320D-02

At iterate   77    f=  7.04612D+01    |proj g|=  6.30766D-02

At iterate   78    f=  7.04611D+01    |proj g|=  6.77607D-02

At iterate   79    f=  7.04610D+01    |proj g|=  3.41741D-02

At iterate   80    f=  7.04609D+01    |proj g|=  4.23538D-02

At iterate   81    f=  7.04608D+01    |proj g|=  4.17224D-02

At iterate   82    f=  7.04607D+01    |proj g|=  5.44741D-02

At iterate   83    f=  7.04606D+01    |proj g|=  3.79311D-02

At iterate   84    f=  7.04606D+01    |proj g|=  5.52782D-02

At iterate   85    f=  7.04606D+01    |proj g|=  6.50172D-02

At iterate   86    f=  7.04605D+01    |proj g|=  3.32194D-02

At iterate   87    f=  7.04605D+01    |proj g|=  2.27349D-02

At iterate   88    f=  7.04604D+01    |proj g|=  2.19097D-02

At iterate   89    f=  7.04604D+01    |proj g|=  2.16217D-02

At iterate   90    f=  7.04604D+01    |proj g|=  4.36989D-02

At iterate   91    f=  7.04603D+01    |proj g|=  2.57841D-02

At iterate   92    f=  7.04603D+01    |proj g|=  1.97333D-02

At iterate   93    f=  7.04603D+01    |proj g|=  1.44884D-02

At iterate   94    f=  7.04603D+01    |proj g|=  3.70448D-02

At iterate   95    f=  7.04603D+01    |proj g|=  7.86955D-03
```

```
At iterate   96    f=  7.04603D+01    |proj g|=  4.72830D-03

At iterate   97    f=  7.04603D+01    |proj g|=  1.08568D-02

At iterate   98    f=  7.04603D+01    |proj g|=  8.33104D-03

At iterate   99    f=  7.04603D+01    |proj g|=  4.69950D-03

At iterate  100    f=  7.04603D+01    |proj g|=  8.87140D-03

At iterate  101    f=  7.04603D+01    |proj g|=  1.19526D-02

At iterate  102    f=  7.04603D+01    |proj g|=  1.03724D-02

At iterate  103    f=  7.04603D+01    |proj g|=  1.31444D-02

At iterate  104    f=  7.04603D+01    |proj g|=  5.07432D-03

At iterate  105    f=  7.04603D+01    |proj g|=  3.44230D-03

At iterate  106    f=  7.04603D+01    |proj g|=  4.56198D-03

At iterate  107    f=  7.04603D+01    |proj g|=  5.09659D-03

At iterate  108    f=  7.04603D+01    |proj g|=  6.83621D-03

At iterate  109    f=  7.04603D+01    |proj g|=  6.24908D-03

At iterate  110    f=  7.04603D+01    |proj g|=  4.14045D-03

At iterate  111    f=  7.04603D+01    |proj g|=  1.05487D-02

At iterate  112    f=  7.04603D+01    |proj g|=  4.58334D-03

At iterate  113    f=  7.04603D+01    |proj g|=  1.51636D-03

           * * *

Tit   = total number of iterations
Tnf   = total number of function evaluations
Tnint = total number of segments explored during Cauchy searches
Skip  = number of BFGS updates skipped
Nact  = number of active bounds at final generalized Cauchy point
Projg = norm of the final projected gradient
F     = final function value

           * * *

   N    Tit     Tnf  Tnint  Skip  Nact     Projg        F
   21    113    129      1     0     0   1.516D-03   7.046D+01
  F =   70.460290012058394

CONVERGENCE: REL_REDUCTION_OF_F_<=_FACTR*EPSMCH
```

Figure



Gradient Descent

Recall in Linear Regression we are trying to find the line

$$y = X\beta$$

that minimizes the sum of square distances between the predicted `y` and the `y` we observed in our dataset:

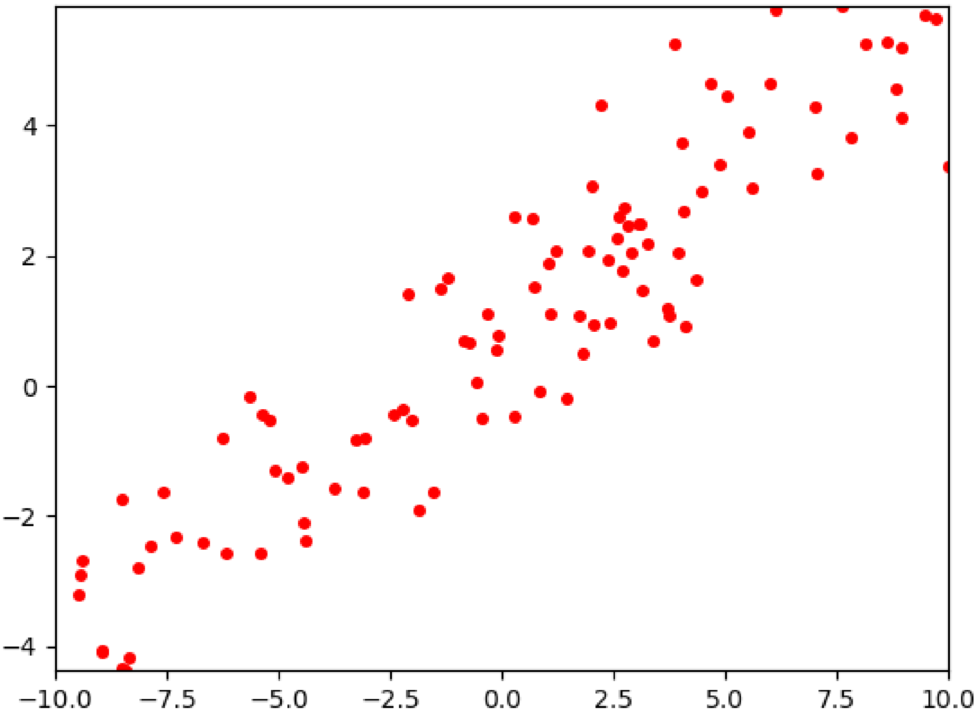
$$\mathcal{L}(\beta) = \|\mathbf{y} - X\beta\|^2$$

We were able to find a global minimum for this loss function, but we will try to apply gradient descent to find the same solution.

a) Implement the `loss` function to complete the code and plot the loss as a function of beta.

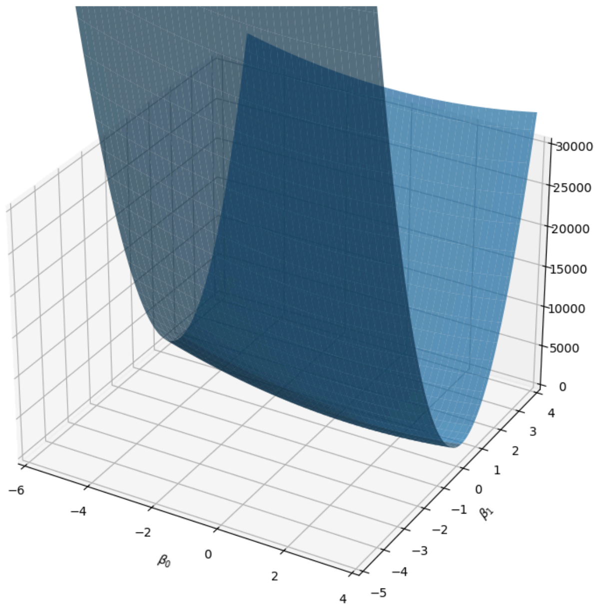
In [ ]:
```
%matplotlib widget
from mpl_toolkits import mplot3d
import numpy as np
import matplotlib.pyplot as plt

beta = np.array([1, .5])
xlin = -10.0 + 20.0 * np.random.random(100)
X = np.column_stack([np.ones((len(xlin), 1)), xlin])
y = beta[0] + (beta[1] * xlin) + np.random.randn(100)

fig, ax = plt.subplots()
ax.plot(xlin, y, 'ro', markersize=4)
ax.set_xlim(-10, 10)
ax.set_ylim(min(y), max(y))
plt.show()
```

Figure



In [ ]:
```
b0 = np.arange(-5, 4, 0.1)
b1 = np.arange(-5, 4, 0.1)
b0, b1 = np.meshgrid(b0, b1)

def loss(X, y, beta):
    return np.sum((y - X @ beta) ** 2)

def get_cost(B0, B1):
    res = []
    for b0, b1 in zip(B0, B1):
        line = []
        for i in range(len(b0)):
            beta = np.array([b0[i], b1[i]])
            line.append(loss(X, y, beta))
        res.append(line)
    return np.array(res)

cost = get_cost(b0, b1)

# create figure
fig = plt.figure(figsize=(14, 9))
ax = plt.axes(projection='3d')
ax.set_xlim(-6, 4)
ax.set_xlabel(r'$\beta_0$')
ax.set_ylabel(r'$\beta_1$')
ax.set_ylim(-5, 4)
ax.set_zlim(0, 30000)

# create plot
ax.plot_surface(b0, b1, cost, alpha=.7)

# show plot
plt.show()
```

Figure



Since the loss is

$$\mathcal{L}(\beta) = \|\mathbf{y} - X\beta\|^2 = \beta^T X^T X\beta - 2\beta^T X^T \mathbf{y} + \mathbf{y}^T \mathbf{y}$$

the gradient would be

$$\nabla_\beta \mathcal{L}(\beta) = 2X^T X\beta - 2X^T \mathbf{y}$$

b) Implement the `gradient` function below and complete the gradient descent algorithm.

In [ ]:
```
import numpy as np
from PIL import Image as im
import matplotlib.pyplot as plt

TEMPFILE = "temp.png"

def snap(betas, losses):
    # Creating figure
    fig = plt.figure(figsize=(14, 9))
    ax = plt.axes(projection='3d')
    ax.view_init(20, -20)
    ax.set_xlim(-5, 4)
    ax.set_xlabel(r'$\beta_0$')
    ax.set_ylabel(r'$\beta_1$')
    ax.set_ylim(-5, 4)
    ax.set_zlim(0, 30000)

    # Creating plot
    ax.plot_surface(b0, b1, cost, color='b', alpha=.7)
    ax.plot(np.array(betas)[:,0], np.array(betas)[:,1], losses, 'o-', c='r', markersize=
    fig.savefig(TEMPFILE)
    plt.close()
    return im.fromarray(np.asarray(im.open(TEMPFILE)))


def gradient(X, y, beta):
    return 2 * X.T @ X @ beta - 2 * X.T @ y
```

```
def gradient_descent(X, y, beta_hat, learning_rate, epochs, images):
    losses = [loss(X, y, beta_hat)]
    betas = [beta_hat]

    for _ in range(epochs):
        images.append(snap(betas, losses))
        beta_hat = beta_hat - learning_rate * gradient(X, y, beta_hat)

        losses.append(loss(X, y, beta_hat))
        betas.append(beta_hat)

    return np.array(betas), np.array(losses)


beta_start = np.array([-5, -2])
learning_rate = 0.0002  # try .0005
# learning_rate = 0.0005  # too large
images = []
betas, losses = gradient_descent(X, y, beta_start, learning_rate, 10, images)

images[0].save(
    'gd.gif',
    optimize=False,
    save_all=True,
    append_images=images[1:],
    loop=0,
    duration=500
)
```

c) Use the code above to create an animation of the linear model learned at every epoch.

In [ ]:
```
def snap_model(beta):
    xplot = np.linspace(-10, 10, 50)
    yestplot = beta[0] + beta[1] * xplot
    fig, ax = plt.subplots()
    ax.plot(xplot, yestplot, 'b-', lw=2)
    ax.plot(xlin, y, 'ro', markersize=4)
    ax.set_xlim(-10, 10)
    ax.set_ylim(min(y), max(y))
    fig.savefig(TEMPFILE)
    plt.close()
    return im.fromarray(np.asarray(im.open(TEMPFILE)))


def gradient_descent(X, y, beta_hat, learning_rate, epochs, images):
    losses = [loss(X, y, beta_hat)]
    betas = [beta_hat]

    for _ in range(epochs):
        images.append(snap_model(beta_hat))
        beta_hat = beta_hat - learning_rate * gradient(X, y, beta_hat)

        losses.append(loss(X, y, beta_hat))
        betas.append(beta_hat)

    return np.array(betas), np.array(losses)


images = []
betas, losses = gradient_descent(X, y, beta_start, learning_rate, 100, images)

images[0].save(
    'model.gif',
    optimize=False,
```

```
    save_all=True,
    append_images=images[1:],
    loop=0,
    duration=200
)
```

In logistic regression, the loss is the negative log-likelihood

$$l(\beta) = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\sigma(x_i\beta)) + (1-y_i)\log(1-\sigma(x_i\beta))$$

the gradient of which is:

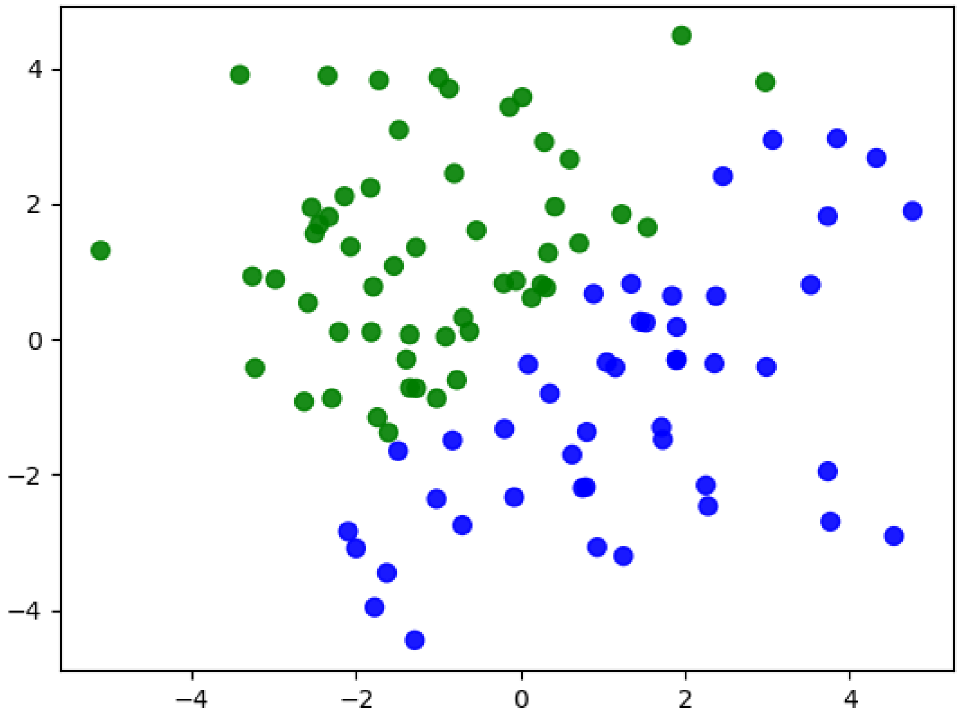
$$\nabla_\beta l(\beta) = -\frac{1}{N}\sum_{i=1}^{N} x_i(y_i - \sigma(x_i\beta))$$

d) Plot the loss as a function of beta.

In [ ]:

```
%matplotlib widget
from mpl_toolkits import mplot3d
import numpy as np
import matplotlib.pyplot as plt
import sklearn.datasets as datasets

centers = [[0, 0]]
t, _ = datasets.make_blobs(n_samples=100, centers=centers, cluster_std=2, random_state=(

# LINE
def generate_line_data():
    # create some space between the classes
    X = t
    Y = np.array([1 if x[0] - x[1] >= 0 else 0 for x in X])
    return X, Y

X, y = generate_line_data()

cs = np.array([x for x in 'gb'])
fig, ax = plt.subplots()
ax.scatter(X[:, 0], X[:, 1], color=cs[y].tolist(), s=50, alpha=0.9)
plt.show()
```

Figure

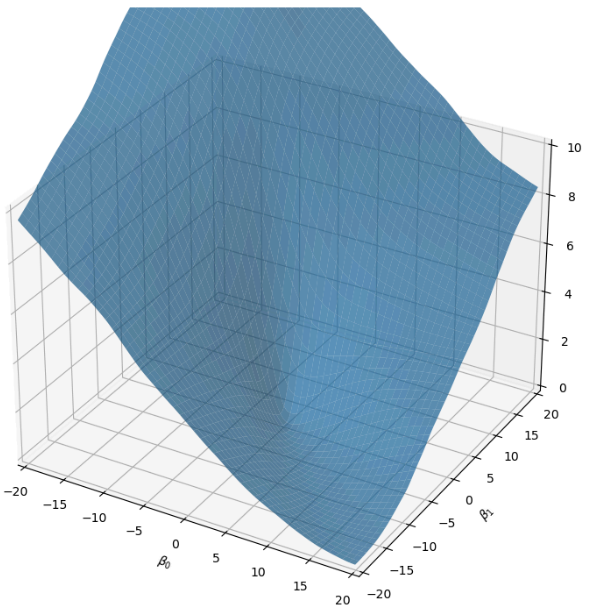


In [ ]:

```
b0 = np.arange(-20, 20, 0.1)
b1 = np.arange(-20, 20, 0.1)
b0, b1 = np.meshgrid(b0, b1)


def sigmoid(x):
    e = np.exp(x)
    return e / (1 + e)


def loss(X, y, beta):
    prob = sigmoid(X @ beta)
    epsilon = 1e-9
    return -np.mean(y * np.log(prob+epsilon) + (1-y) * np.log(1-prob+epsilon))


def get_cost(B0, B1):
    res = []
    for b0, b1 in zip(B0, B1):
        line = []
        for i in range(len(b0)):
            beta = np.array([b0[i], b1[i]])
            line.append(loss(X, y, beta))
        res.append(line)
    return np.array(res)

cost = get_cost(b0, b1)


# create figure
fig = plt.figure(figsize=(14, 9))
ax = plt.axes(projection='3d')
ax.set_xlim(-20, 20)
ax.set_xlabel(r'$\beta_0$')
ax.set_ylabel(r'$\beta_1$')
```

```
ax.set_ylim(-20, 20)
ax.set_zlim(0, 10)

# create plot
ax.plot_surface(b0, b1, cost, alpha=.7)

# show plot
plt.show()
```

Figure



e) Plot the loss at each iteration of the gradient descent algorithm.

In [ ]:
```
import numpy as np
from PIL import Image as im
import matplotlib.pyplot as plt


TEMPFILE = "temp.png"


def snap(betas, losses):
    # Creating figure
    fig = plt.figure(figsize=(14, 9))
    ax = plt.axes(projection='3d')
    ax.view_init(10, 10)
    ax.set_xlabel(r'$\beta_0$')
    ax.set_ylabel(r'$\beta_1$')
    ax.set_ylim(-20, 20)
    ax.set_zlim(0, 10)

    # Creating plot
    ax.plot_surface(b0, b1, cost, color='b', alpha=.7)
    ax.plot(np.array(betas)[:,0], np.array(betas)[:,1], losses, 'o-', c='r', markersize
    fig.savefig(TEMPFILE)
    plt.close()
    return im.fromarray(np.asarray(im.open(TEMPFILE)))


def gradient(X, y, beta):
```

```python
    prob = sigmoid(X @ beta)
    return -X.T @ (y-prob) / len(y)


def gradient_descent(X, y, beta_hat, learning_rate, epochs, images):
    losses = [loss(X, y, beta_hat)]
    betas = [beta_hat]

    for _ in range(epochs):
        images.append(snap(betas, losses))
        beta_hat = beta_hat - learning_rate * gradient(X, y, beta_hat)

        losses.append(loss(X, y, beta_hat))
        betas.append(beta_hat)

    return np.array(betas), np.array(losses)


beta_start = np.array([-5, -2])
learning_rate = 0.1
images = []
betas, losses = gradient_descent(X, y, beta_start, learning_rate, 10, images)

images[0].save(
    'gd_logit.gif',
    optimize=False,
    save_all=True,
    append_images=images[1:],
    loop=0,
    duration=500
)
```

f) Create an animation of the logistic regression fit at every epoch.

In [ ]:
```python
def snap_model(beta):
    xplot = np.linspace(-10, 10, 50)
    yestplot = beta[1] + beta[0] * xplot
    fig, ax = plt.subplots()
    ax.plot(xplot, yestplot, 'b-', lw=2)
    ax.scatter(X[:, 0], X[:, 1], color=cs[y].tolist())
    ax.set_xlim(-5, 5)
    ax.set_ylim(-5, 5)
    fig.savefig(TEMPFILE)
    plt.close()
    return im.fromarray(np.asarray(im.open(TEMPFILE)))

def gradient_descent(X, y, beta_hat, learning_rate, epochs, images):
    losses = [loss(X, y, beta_hat)]
    betas = [beta_hat]

    for _ in range(epochs):
        images.append(snap_model(beta_hat))

        beta_hat = beta_hat - learning_rate * gradient(X, y, beta_hat)

        losses.append(loss(X, y, beta_hat))
        betas.append(beta_hat)

    return np.array(betas), np.array(losses)

images = []
betas, losses = gradient_descent(X, y, beta_start, learning_rate, 120, images)

images[0].save(
    'model_logit.gif',
    optimize=False,
```

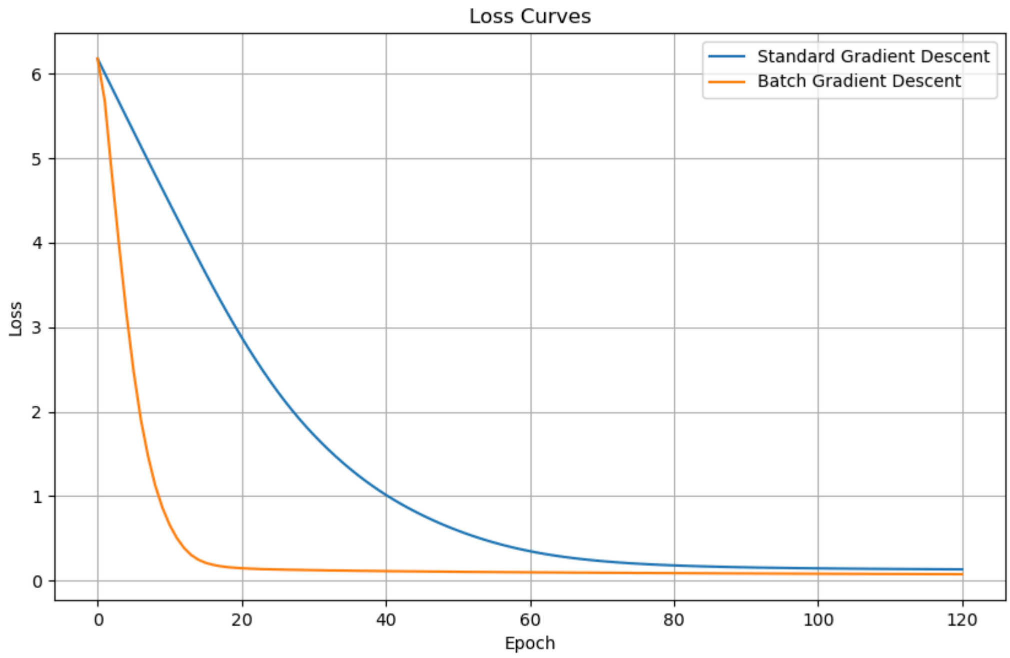
```
    save_all=True,
    append_images=images[1:],
    loop=0,
    duration=200
)
```

g) Modify the above code to evaluate the gradient on a random batch of the data. Overlay the true loss curve and the approximation of the loss in your animation.

In [ ]:
```
def gradient_descent(X, y, beta_hat, learning_rate, epochs, images):
    losses = [loss(X, y, beta_hat)]
    betas = [beta_hat]

    for _ in range(epochs):
        images.append(snap_model(beta_hat))

        beta_hat = beta_hat - learning_rate * gradient(X, y, beta_hat)

        losses.append(loss(X, y, beta_hat))
        betas.append(beta_hat)

    return np.array(betas), np.array(losses)


def batch_gradient_descent(X, y, beta_hat, learning_rate, epochs, batch_size, images):
    n_samples = len(y)
    losses = [loss(X, y, beta_hat)]
    betas = [beta_hat]

    for _ in range(epochs):
        images.append(snap_model(beta_hat))

        permuted_indices = np.random.permutation(n_samples)
        batch_losses = []

        for i in range(0, n_samples, batch_size):
            batch_indices = permuted_indices[i:i+batch_size]
            X_batch = X[batch_indices]
            y_batch = y[batch_indices]

            beta_hat = beta_hat - learning_rate * gradient(X_batch, y_batch, beta_hat)
            batch_losses.append(loss(X_batch, y_batch, beta_hat))

        losses.append(np.mean(batch_losses))
        betas.append(beta_hat)

    return np.array(betas), np.array(losses)


beta_start = np.array([-5, -2])
learning_rate = 0.1
epochs = 120
batch_size = 20
images = []
b_images = []
betas, losses = gradient_descent(X, y, beta_start, learning_rate, epochs, images)
b_betas, b_losses = batch_gradient_descent(X, y, beta_start, learning_rate, epochs, bat

b_images[0].save(
    'model_logit_batch.gif',
    optimize=False,
    save_all=True,
    append_images=b_images[1:],
    loop=0,
    duration=200
)
```

```
plt.figure(figsize=(10, 6))
plt.plot(range(epochs+1), losses, label='Standard Gradient Descent')
plt.plot(range(epochs+1), b_losses, label='Batch Gradient Descent')
plt.title('Loss Curves')
plt.xlabel('Epoch')
plt.ylabel('Loss')
plt.legend()
plt.grid(True)
plt.show()
```

Figure



h) Below is a sandbox where you can get intuition about how to tune the parameters:

In [ ]:
```
import numpy as np
from PIL import Image as im
import matplotlib.pyplot as plt

TEMPFILE = "temp.png"

def snap(x, y, pts, losses, grad):
    fig = plt.figure(figsize=(14, 9))
    ax = plt.axes(projection='3d')
    ax.view_init(20, -20)
    ax.plot_surface(x, y, loss(np.array([x, y])), color='r', alpha=.4)
    ax.plot(np.array(pts)[:,0], np.array(pts)[:,1], losses, 'o-', c='b', markersize=10,
    ax.plot(np.array(pts)[-1,0], np.array(pts)[-1,1], -1, 'o-', c='b', alpha=.5, marker

    # Plot Gradient Vector
    X, Y, Z = [pts[-1][0]], [pts[-1][1]], [-1]
    U, V, W = [-grad[0]], [-grad[1]], [0]
    ax.quiver(X, Y, Z, U, V, W, color='g')
    fig.savefig(TEMPFILE)
    plt.close()
    return im.fromarray(np.asarray(im.open(TEMPFILE)))

def loss(x):
    return np.sin(sum(x**2)) # changeable

def gradient(x):
    return 2 * x * np.cos(sum(x**2)) # changeable
```

```python
def gradient_descent(x, y, init, learning_rate, epochs):
    images, losses, pts = [], [loss(init)], [init]
    for _ in range(epochs):
        grad = gradient(init)
        images.append(snap(x, y, pts, losses, grad))
        init = init - learning_rate * grad
        losses.append(loss(init))
        pts.append(init)
    return images

init = np.array([-.5, -.5]) # changeable
learning_rate = 1.394 # changeable
x, y = np.meshgrid(np.arange(-2, 2, 0.1), np.arange(-2, 2, 0.1)) # changeable
images = gradient_descent(x, y, init, learning_rate, 12)

images[0].save(
    'gradient_descent.gif',
    optimize=False,
    save_all=True,
    append_images=images[1:],
    loop=0,
    duration=500
)
```