Exercise 1

First step will be to generate the data.

- x_1 and x_2 come from the normal distribution N(0,1)
- y is calculated based on the formula $y = 3 + 2x_1 + x_2 + x_1 x_2$

Noise is also added to the data.

Now, ignoring the formula we will try to calculate the parameters of the model utilizing the data and the LSE criterion:

$$\theta = (X^T X) X^T y$$

Matrix X looks like this:

```
array([[ 1.        ,  1.35441089,  0.87832556,  1.18961371],
       [ 1.        , -1.80757007,  0.26021866, -0.47036346],
       [ 1.        , -0.13069749,  0.82020266, -0.10719843],
       [ 1.        ,  0.32995238, -1.19347151, -0.39378877],
       [ 1.        , -0.7938124 , -1.18317656,  0.93922023],
       [ 1.        ,  0.94425241, -1.37878186, -1.3019181 ],
       [ 1.        , -1.28896619,  0.6034808 , -0.77786635],
       [ 1.        , -0.80198341,  1.70286324, -1.36566807],
```

The estimation of theta vector is:

```
array([2.99713987, 1.97799976, 0.99592521, 0.98525395])
```

We can observe that theta values are close to the real theta values that generated the data.

Exercise 2

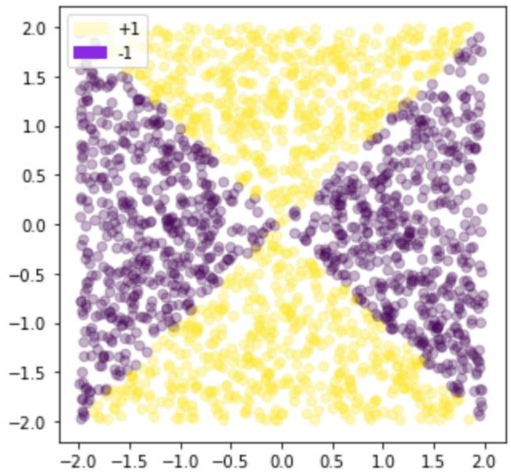
First step will be to generate the data.

- x_1 and x_2 come from the uniform distribution, with range (-2, 2)
- y is calculated based on the formula $y = x_2^2 - x_1^2$

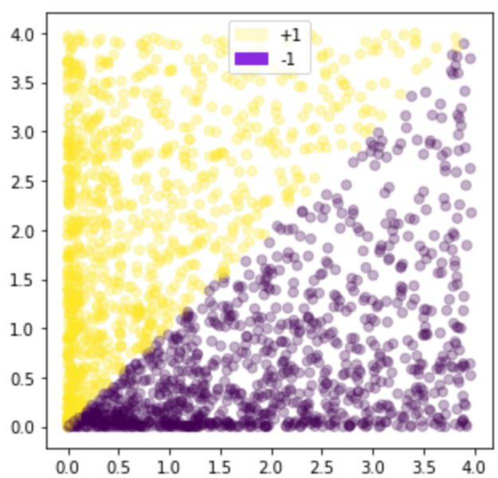
Next, we classify the data. Those who lie on the positive side of the curve are assigned to +1 class and the rest to -1 class.

```
for i in range(len(X)):
    rule = (X[i][1])**2 - (X[i][0])**2
    y.append(1) if rule>=0 else y.append(-1)
```

Data plotted look like this:



And after we transform the problem to linear, data look like this:



Now we will use the MSE criterion $\theta = (X^T X) X^T y$, to calculate the theta vector or the parameters of the model.

Theta vector is:

```
array([ 0.08284334, -0.49356928,  0.44659933])
```

Exercise 3

The non-linear problem is:

$$y = 3x_1^2 + 4x_2^2 + 5x_3^2 + 7x_1x_2 + x_1x_3 + 4x_2x_3 - 2x_1 - 3x_2 - 5x_3 + \eta$$

The problem is defined in three dimensions: x = [x_1, x_2, x_3] $^T \in R^3$

The function ϕ that transforms the problem to a space where it is linear is:

$$\varphi(x) = \begin{bmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \varphi_3(x) \\ \varphi_4(x) \\ \varphi_5(x) \\ \varphi_6(x) \\ \varphi_7(x) \\ \varphi_8(x) \\ \varphi_9(x) \end{bmatrix} = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_3^2 \\ x_1x_2 \\ x_1x_3 \\ x_2x_3 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}$$ thus, the problem can be expressed as:

$$y = 3\varphi_1(x) + 4\varphi_2(x) + 5\varphi_3(x) + 7\varphi_4(x) + \varphi_5(x) + 4\varphi_6(x) - 2\varphi_7(x) - 3\varphi_8(x) - 5\varphi_9(x) + \eta$$

The dimensions of the transformed problem are nine: $\varphi(x) \in R^9$

Exercise 4

The non-linear problem is:

$$x_1^2 + 3x_2^2 + 6x_3^2 + x_1x_2 + x_2x_3 > (<)3 \rightarrow x \in \omega_1(w_2)$$

The problem is defined in three dimensions: x = [x_1, x_2, x_3] $^T \in R^3$

The function ϕ that transforms the problem to a space where it is linear is:

$$\varphi(x) = \begin{bmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \varphi_3(x) \\ \varphi_4(x) \\ \varphi_5(x) \end{bmatrix} = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_3^2 \\ x_1x_2 \\ x_2x_3 \end{bmatrix}$$ thus, the problem can be expressed as:

$$\varphi_1(x) + 3\varphi_2(x) + 6\varphi_3(x) + \varphi_4(x) + \varphi_5(x) = 0$$

The dimensions of the transformed problem are five: $\varphi(x) \in R^5$

Exercise 5

Our data in matrices (y for the class and X for the data points) are:

$$y = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 0 & 0.5 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix}$$

Taking advantage of the least squares error criterion $\hat{\theta} = (X^T X)^{-1} X^T y$, we will find the linear model that best describes the problem.

First, we need to prove that the inverse matrix of $X^T X$ exists.

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 0 & 0.5 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 0 & 0.5 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0.5 \\ 0 & 4 & 0 \\ 0.5 & 0 & 4.25 \end{bmatrix}$$

Thus $(X^T X)^{-1} = \begin{bmatrix} 0.2024 & 0 & -0.238 \\ 0 & 0.25 & 0 \\ -0.238 & 0 & 0.238 \end{bmatrix}$

In conclusion, $\hat{\theta} = (X^T X)^{-1} X^T y = \begin{bmatrix} 0.1905 \\ 1 \\ 0.0952 \end{bmatrix}$

Linear classifier is: $\varphi_1(x_1) + 0.0952\, \varphi_2(x_2) + 0.1905 = 0$

Exercise 6

- Sum rule: $P(x) = \sum_{y \in Y} P(x,y)$

$$\sum_{y \in Y} P(x,y) = \sum_{j=1}^{n_y} P(x, y_j) = \sum_{j=1}^{n_y} \frac{n_{ij}}{n} = \frac{\sum_{j=1}^{n_y} n_{ij}}{n}$$

But $\sum_{j=1}^{n_y} n_{ij} = n_i^x$ thus,

$$\sum_{y \in Y} P(x,y) = \frac{n_i^x}{n} = P(x)$$

- Product rule: $P(x,y) = P(x|y)P(y)$

$P(x|y)$ can be expressed as "the times that x occurred; while y occurred", or $P(x|y) = \frac{n_{ij}}{n_j}$

$$P(x|y) = \frac{n_{ij}}{n_j} = \frac{\frac{n_{ij}}{n}}{\frac{n_j}{n}} = \frac{P(x,y)}{P(y)}$$

- Bayes rule: $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$

$$\frac{P(x|y)P(y)}{P(x)} = \frac{\frac{n_{ij}}{n_j} \cdot \frac{n_j}{n}}{\frac{n_i}{n}} = \frac{n_{ij}}{n_i} = P(y|x)$$