Assignment 1

1) A binary classifier decides whether data is from one of two classes, labeled 1 and -1. In this problem the data is described by two features x_1 and x_2 and the classification decision is made as follows. The class with label 1 is decided if $x_1a_1 + x_2a_2 > b$ while the class with label -1 is decided when $x_1a_1 + x_2a_2 < b$. Here a_1, a_2, and b are given real numbers.

   a) The *decision boundary* is the set of $\{x_1, x_2\}$ that satisfy $x_1a_1 + x_2a_2 = b$. Thus, we may assign the label to the data using the sign of $y = x_1a_1 + x_2a_2 - b$ since label 1 is decided if $y > 0$ and label -1 is decided if $y < 0$. That is, the label may be obtained as sign$\{y\}$. Express y as an inner product of a vector $\boldsymbol{x}$ containing the features and $\boldsymbol{w}$ containing weights, that is, write $y = \boldsymbol{x}^T\boldsymbol{w}$.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ -1 \end{bmatrix} \quad w = \begin{bmatrix} a_1 \\ a_2 \\ b \end{bmatrix} \quad y = \begin{bmatrix} x_1 & x_2 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ b \end{bmatrix}$$

   b) Let x_2 be the vertical axis and x_1 be the horizontal axis. Show that the decision boundary $y = 0$ is a straight line. Find the slope and intercept with the vertical axis as a function of a_1, a_2, b.

$$0 = x_1a_1 + x_2a_2 - b$$
$$x_2a_2 = b - x_1a_1$$
$$x_2 = \frac{b}{a_2} - \frac{x_1a_1}{a_2}$$

x_2 is vertical axis which is the y-axis.
Formula for a straight line → $y = mx + b$
The equation above matches the formula for a straight line.

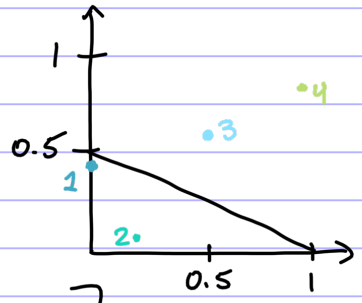
m = slope = $-\frac{a_1}{a_2}$	b = vertical intercept = $\frac{b}{a_2}$

   c) You classify n data samples using sign$\{\boldsymbol{y}\}$ where $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \boldsymbol{X}\boldsymbol{w}$. Suppose $n = 4$ and the features for the 4 data samples are $1 : (0, 0.4), 2 : (0.2, 0.1), 3 : (0.5, 0.6), 4 : (0.9, 0.8)$. Write out the matrix $\boldsymbol{X}$.

$$X = \begin{bmatrix} 0 & 0.4 & -1 \\ 0.2 & 0.1 & -1 \\ 0.5 & 0.6 & -1 \\ 0.9 & 0.8 & -1 \end{bmatrix}$$

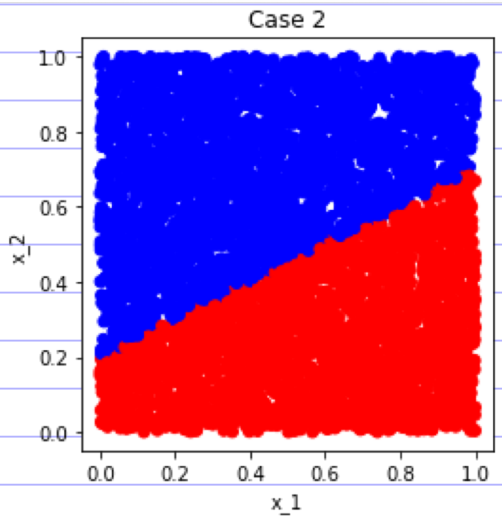
d) Suppose $a_1 = 1$, $a_2 = 2$, and $b = 1$. Sketch the decision boundary in the x_1-x_2 plane assuming x_2 is the vertical axis and x_1 is the horizontal axis. Graph the four data samples from the previous part and classify them.

$$y = x_1 + 2x_2 - 1$$
$$0 = x_1 + 2x_2 - 1$$
$$2x_2 = 1 - x_1$$
$$x_2 = \frac{1}{2} - \frac{x_1}{2}$$



[1 and 2 are -1
3 and 4 are 1]

e) Download and run the linear classifier script. This script classifies 5000 examples of (randomly generated) data consisting of two features using the linear classifier. Save the figure and include it in your submission. Describe the decision boundary you observe using a sentence.



The boundary is positive and linear with an intercept with the vertical axis at 0.2.

f) Change the classifier weights to $w = [1.6 \quad 2 \quad -1.6]^T$. Rerun the scrip. Include the figure in your pdf file. Briefly describe how the change in the weights changed the decision boundary.

The direction of the linear boundary changed from positive (increasing) to negative (decreasing) and the boundary appears a little steeper.