#Homework 01

Step 1

What is Machine Learning?

Machine Learning (ML) is a very far-reaching subject in computer science with important uses in data science, artificial intelligence, and more. It is a data-oriented approach to train a system into completing a certain task. Regardless of purpose, most machine learning models depend on data being fed into it. For models that predict or classify data into categories, more data used in training usually equates to higher accuracy when utilizing the model on new data. ML uses many features in statistics and computer science to help make more accurate decisions, such as regression models, decision trees, and error analysis functions.

Step 2

```
In [1]: import numpy as np
        import matplotlib.pyplot as plt
        x = np.linspace(0, 10, 100)
        plt.plot(x, np.sin(x), c = "green")
        plt.title("Step 2")
        plt.xlabel("x")
        plt.ylabel("y = sin(x)")
        plt.grid(b = True, which = "Major", axis = "both", c = "gray", alpha = 0.5)
        plt.show()

        import seaborn as sb
        plot2 = sb.lineplot(x = x, y = np.sin(x), color = "green")
        sb.set_style("whitegrid", {"grid.color": "gray", "grid.alpha": "0.5"})
        plot2.set(title = "Step 2", xlabel = "x", ylabel = "y = sin(x)")
```

```
<Figure size 640x480 with 1 Axes>
```

```
Out[1]: [Text(0, 0.5, 'y = sin(x)'), Text(0.5, 0, 'x'), Text(0.5, 1.0, 'Step 2')]
```
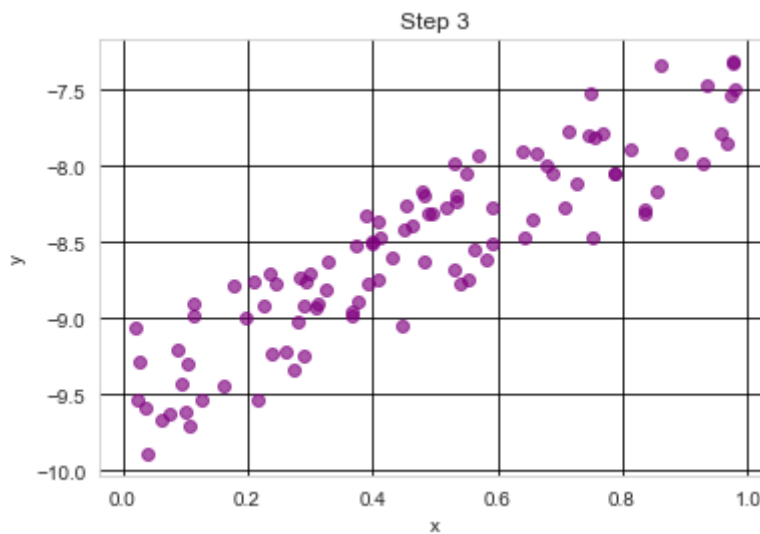
Step 3

```
In [2]: import numpy.random as npr
        import matplotlib.pyplot as plt

        x = npr.random_sample(size = 100)
        z = npr.random_sample(size = 100)
        y = 2*x - 10 + z

        plt.scatter(x = x, y = y, c = "purple", alpha = 0.65)
        plt.grid(b = True, which = "Major", axis = "both", c = "black")
        plt.title("Step 3")
        plt.xlabel("x")
        plt.ylabel("y")
```

Out[2]: Text(0, 0.5, 'y')

Step 4

```
In [3]: f = 50*x - 20*x*y + 10*y + 18 + z
```

The above model is a function of 2 variables (x, y) with z being noise. This model to predict age assumes that x and y have way more impact on the output than noise z. The model takes the general form f = Ax - Bxy + Cy + c + z, with A, B, C, and c being positive constants. This model takes into consideration of all edge cases i.e. combinations of cane present/absent and suitcase present/absent. The constant c = 18 describes the age if neither cane nor suitcase are present. 18 is a good initial age being the age when people just graduate high school. Noise z may result in

younger ages which obviously do not work and do not rely on a cane to walk, or result in ages slightly above 18 corresponding to young adults in college and not yet employed. If only the suitcase is present, then y = 1 and 10 is added, resulting in age centered around 28. This approximately aligns with Ph.D., medical school, and some vocational school graduates who just entered the workforce. Middle-age working people are represented by positive z added to 28. People who have worked for a long time but not retired yet have a predicted age centered around 58, with x = y = 1. Finally, retirees and elders indicate that only x = 1, resulting in a predicted age centered around 68. Of course, coefficients may be tweaked differently for different scenarios, such as adjusting to a specific country's labor laws and retirement laws.