Reading: Class notes: week 3.

Problems:

1. In Example 16 from the class note, we derived the least-squares solution for linear regression in the case of scalar real-valued inputs, $\{x^i : i = 1, 2, \ldots, N\}$, and outputs, $\{y^i : i = 1, 2, \ldots, N\}$, without a bias term. Now, we will extend this to include an additional bias weight, w_0.

The linear regression model with a bias term is now given by:

$$y = wx + w_0$$

where w is the slope (weight) and w_0 is the bias.

1. **Derive the updated least-squares cost function** that includes the bias term w_0. Show that the cost function to minimize is:

$$J(w, w_0) = \frac{1}{2N} \sum_{i=1}^{N} \left(y^i - (wx^i + w_0)\right)^2$$

2. **Compute the partial derivatives** of $J(w, w_0)$ with respect to both w and w_0. Verify that

$$\frac{\partial J}{\partial w} = -\frac{1}{N} \sum_{i=1}^{N} \left(y^i - (wx^i + w_0)\right) x^i$$
$$\frac{\partial J}{\partial w_0} = -\frac{1}{N} \sum_{i=1}^{N} \left(y^i - (wx^i + w_0)\right)$$

3. **Set the partial derivatives to zero** and obtain a system of simultaneous equations. Express the equations as

$$\begin{bmatrix} 1 & * \\ * & * \end{bmatrix} \begin{bmatrix} w_0 \\ w \end{bmatrix} = \begin{bmatrix} * \\ * \end{bmatrix}$$

4. **Solve the system of equations** to find the optimal values of w and w_0. Show that the solution cana be expressed as follows:

$$w = \frac{\sum_{i=1}^{N} (x^i - \hat{x})(y^i - \hat{y})}{\sum_{i=1}^{N} (x^i - \hat{x})^2}, \quad w_0 = \hat{y} - w\hat{x}$$

where

$$\hat{x} = \frac{1}{N} \sum_{i=1}^{N} x^i, \qquad \hat{y} = \frac{1}{N} \sum_{i=1}^{N} y^i$$

[Hint: To solve the equations, it may be useful to first verify the following formula for inverse of a 2×2 matrix.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \qquad (ad - bc) \neq 0$$

]

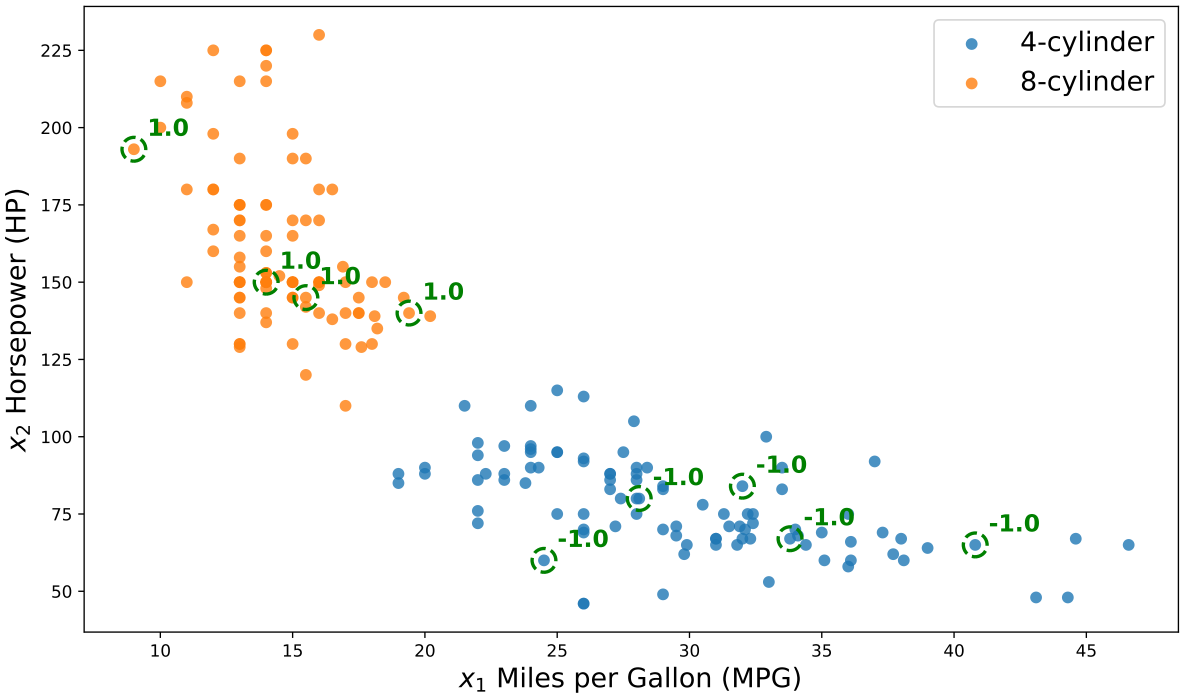

Figure 1: Scatter plot of the dataset. Each point represent a data sample with two features (x-axis and y-axis) and a categroy (color). The categories are represented by colors: blue (4-cylinder engines) and orange (8-cylinder engines).

The following questions are based on a dataset of vehicle characteristics that describes a relationship between miles per gallon (MPG), horsepower (HP), and the number of cylinders in the engine. The dataset and example Python scripts are provided, and can be accessed on GitHub.

2. Data Visualization and Interpretation: Examine the scatter plot (Figure 1) of the dataset:

1. Describe the general features of data points. What might be the reason for such a distribution of data points? How many categories (labels) are there in the dataset? Is it possible to separate the categories?
2. The green circles on data points indicate a few randomly selected samples. Run the Python code and read from terminal output to answer the following questions.
   (a) What is the meaning of the green numbers for these highlighted data points? Note that these numbers are either 1 or -1?
   (b) How many distinct samples N are there in the dataset?
3. Create a mathematical model for this data-set as a single layer perceptron

$$y = \sigma(w^{\mathrm{T}}x)$$

   (a) What are the possible values of y? [Hint: These are related to the green numbers].
   (b) What is the form of x? [Hint: These are related to the the two axes and a bias].
   (c) What is the choice of $\sigma(\cdot)$?

3. A binary classifier has been implemented using PyTorch to classify 4-cylinder and 8-cylinder engines. The decision boundary (separating the two categories) is given by

$$w^\intercal x = \begin{bmatrix} w_0 & w_1 & w_2 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} = 0 \tag{1}$$

1. By inspection or trial and error, find one example of numerical values for w_0, w_1, w_2 that *does not* separate the dataset.
2. By inspection or trial and error, find one example of numerical values for w_0, w_1, w_2 that *does* separate the dataset.
3. Turn in a plot showing the two decision boundaries – with the incorrect and the correct choice.
4. For each choice, compute and report the resulted value of the loss function

$$J(w_0, w_1, w_2) = \frac{1}{2N} \sum_{i=1}^{N} (y^i - (w_1 x_1^i + w_2 x_2^i + w_0))^2$$

4. Least square algorithm: A code has been provided to implement the least squares algorithm

1. Run the code to obtain the optimal decision boundary.
2. Does the optimal weight vector obtained by the least squares algorithm separate the two groups (4-cylinder and 8-cylinder engines)?
3. Turn in a plot showing the optimal decision boundary. Report the numerical value of the optimal weights.
4. Report the numerical value of the loss. Compare with the numerical value of the loss for your two choices in Problem 3.