Week 1 Worksheet — Simple Linear Classifier

Health Mini-Project (No Hidden Nodes)

Name: _____ **Date:** _____

Goal: Design your own linear rule to classify each person as Healthy or Unhealthy using four input features.

Important: There is no single correct answer. Different students may create different, logical rules.

Dataset (Inputs and Outputs)

Use **IDs 1–10** to create your rule ("training") and **IDs 11–12** to test it.

ID	Height (in)	Weight (lb)	Waist (in)	Favorite Color	Your Score	Your Label
1	70	159	32	Blue		
2	65	187	37	Red		
3	72	198	39	Green		
4	63	121	28	Yellow		
5	68	150	31	Black		
6	67	209	41	Blue		
7	71	172	33	Pink		
8	62	181	36	Purple		
9	69	146	31	Orange		
10	73	231	43	White		
11	64	128	28	Green		
12	70	154	33	Red		

Your Task:

1. Assign importance (weights) to each input: Height, Weight, Waist, and Favorite Color. Example: $w_1 = 0.2$, $w_2 = 0.4$, $w_3 = 0.3$, $w_4 = 0.1$, with $w_1 + w_2 + w_3 + w_4 = 1$.

2. Define your scoring rule:

$$\text{Score} = w_1(\text{Height}) + w_2(\text{Weight}) + w_3(\text{Waist}) + w_4(\text{ColorCode})$$

   (Note: Favorite Color can be converted into numbers however you think makes sense.)

3. **Think about the problem of scale (Normalization).**
   Height, Weight, and Waist have very different numeric ranges. The feature with the largest numbers may dominate your Score.

To balance them, normalize each feature using this formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

After normalization, all features are between 0 and 1. Recompute your Score using normalized values and compare with the previous result.

Question: What changed after normalization? Did the relative importance of your features shift?

4. Choose a threshold T. Example: $T = 150$.
5. Classify: If Score $< T \rightarrow$ Healthy, else Unhealthy.
6. Apply your rule to IDs 1–10. Fill in "Your Score" and "Your Label" columns.
7. Test your rule on IDs 11–12.

Reflection:

- Which variable(s) influenced your classification the most? Why?
- How did you decide the threshold value T?
- If you changed one weight, how might your predictions change?

Hint: You can start simple by focusing on two features (like Weight and Waist) and expand later.

Challenge (Optional): Search for the Best Weights with Code

Until now you chose weights by reasoning. Suppose the true labels for IDs 1–10 are known (released after submission). Write code to **search** for weights that maximize training accuracy, then test on IDs 11–12.

What to do

1. Normalize each feature (min–max over the training set).
2. Search weights w_1, w_2, w_3, w_4 with $w_i \geq 0$ and $\sum w_i = 1$ (grid step 0.1).
3. For each weight set and threshold T (also grid), compute Score and accuracy on IDs 1–10.
4. Keep the best (w, T). Report training accuracy and test predictions for IDs 11–12.