CSCI 303 Introduction to Data Science - Midterm Concept Check

1. Explain what would be represented by: X, y, y_hat

2. What is the difference between supervised vs. unsupervised machine learning?

3. We have discussed 8 different steps in the Regression Workflow last week … Please list them below in order and provide a reason we would want to do each step.

Workflow Step	Why	How: Pseudocode/Code Sample
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		

4. What are the two different reasons / places we would use visualizations in our workflow?

5. Which one of the below would represent an overfitting, underfitting, or balanced situation (please write below the images)?