Dr. Denton Bobeldyk

CIS 365 Artificial Intelligence

KNN Classification

Week in Review

Blackboard Check-in

Delivery Methods

Lecture

Videos

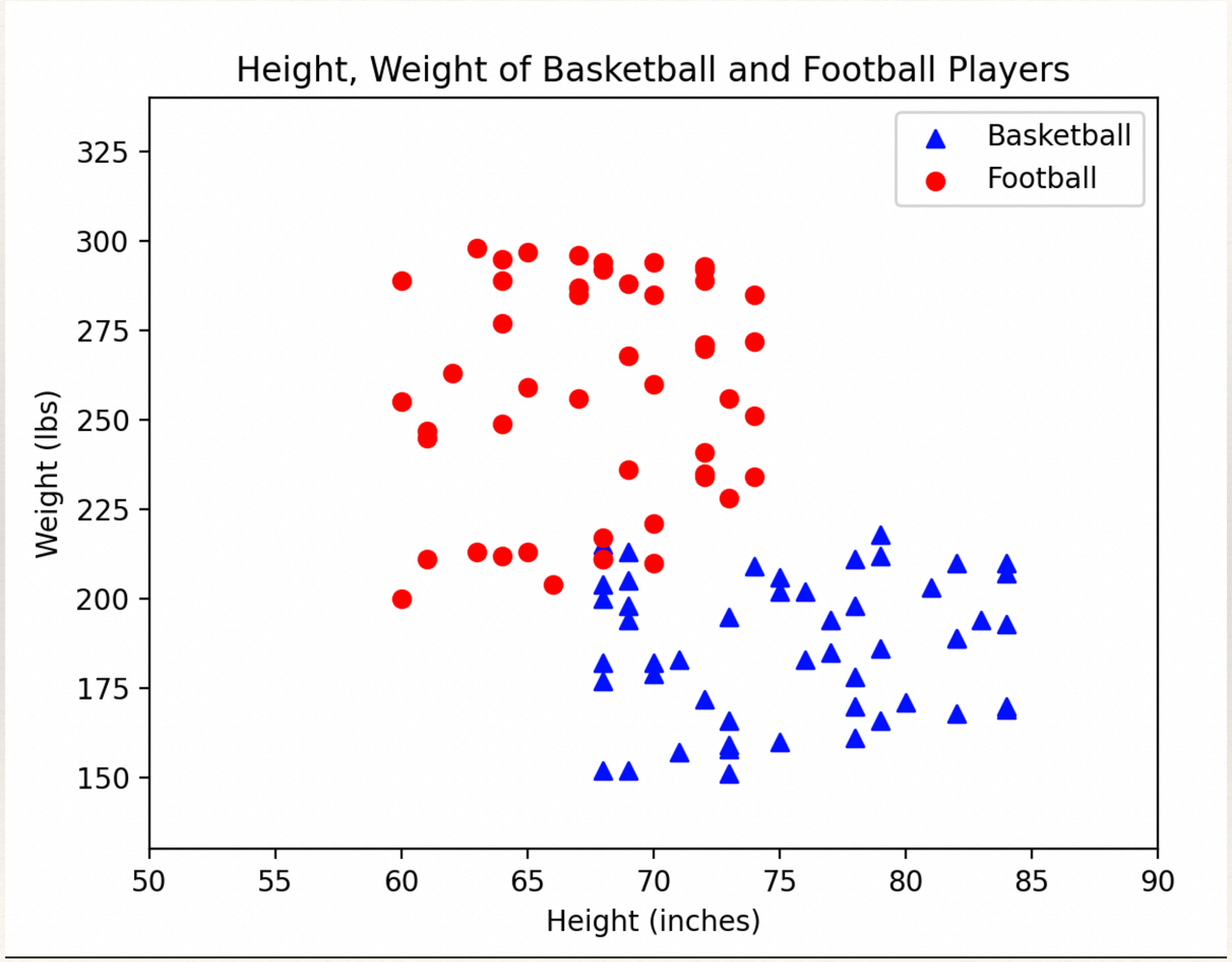
Lab Time

Small Groups

Classification

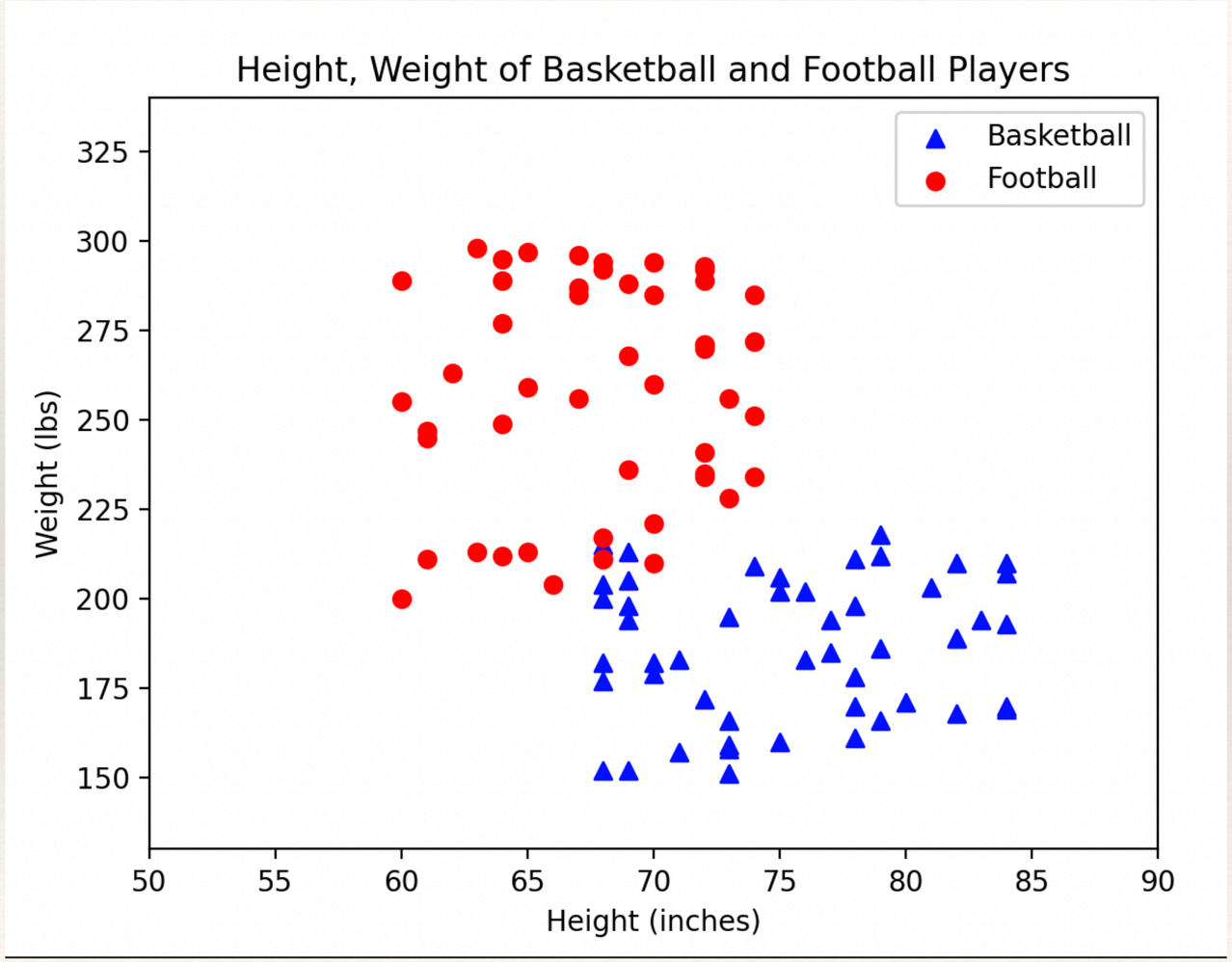
- Assign a data point to one of a predefined set of categories or classes
- For example:
  - Determining if an email is spam or not spam
  - Categorizing an image into types (e.g., cat, dog, horse)

Classification



Given the weight and height of an athlete, we can classify them as either a Basketball player or Football player

Feature Vector



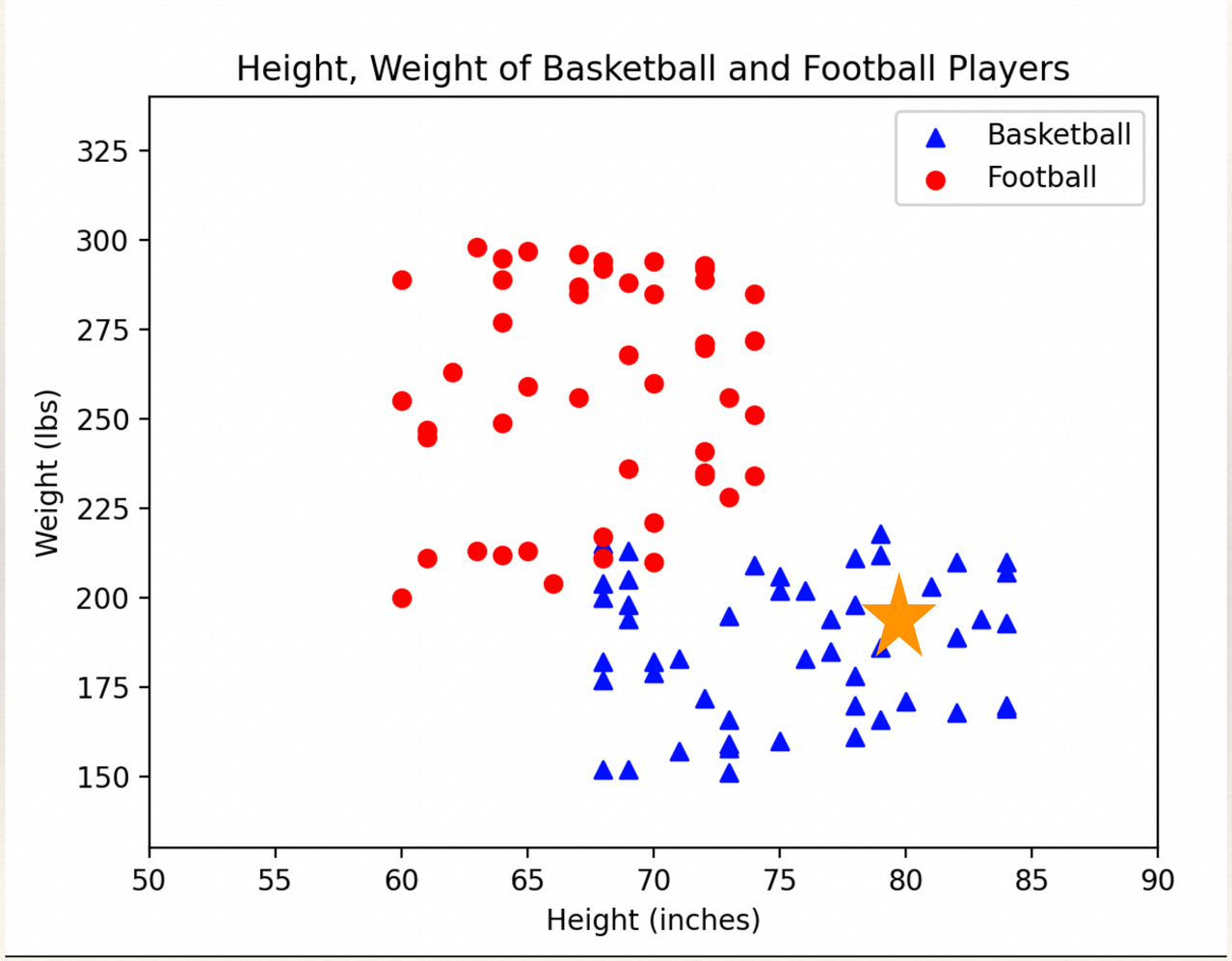
The attributes of a data point are typically called features and a feature vector is simply a collection of those attributes for each item

In the example shown on the left, we have plotted the two features for each athlete

Classification Methods

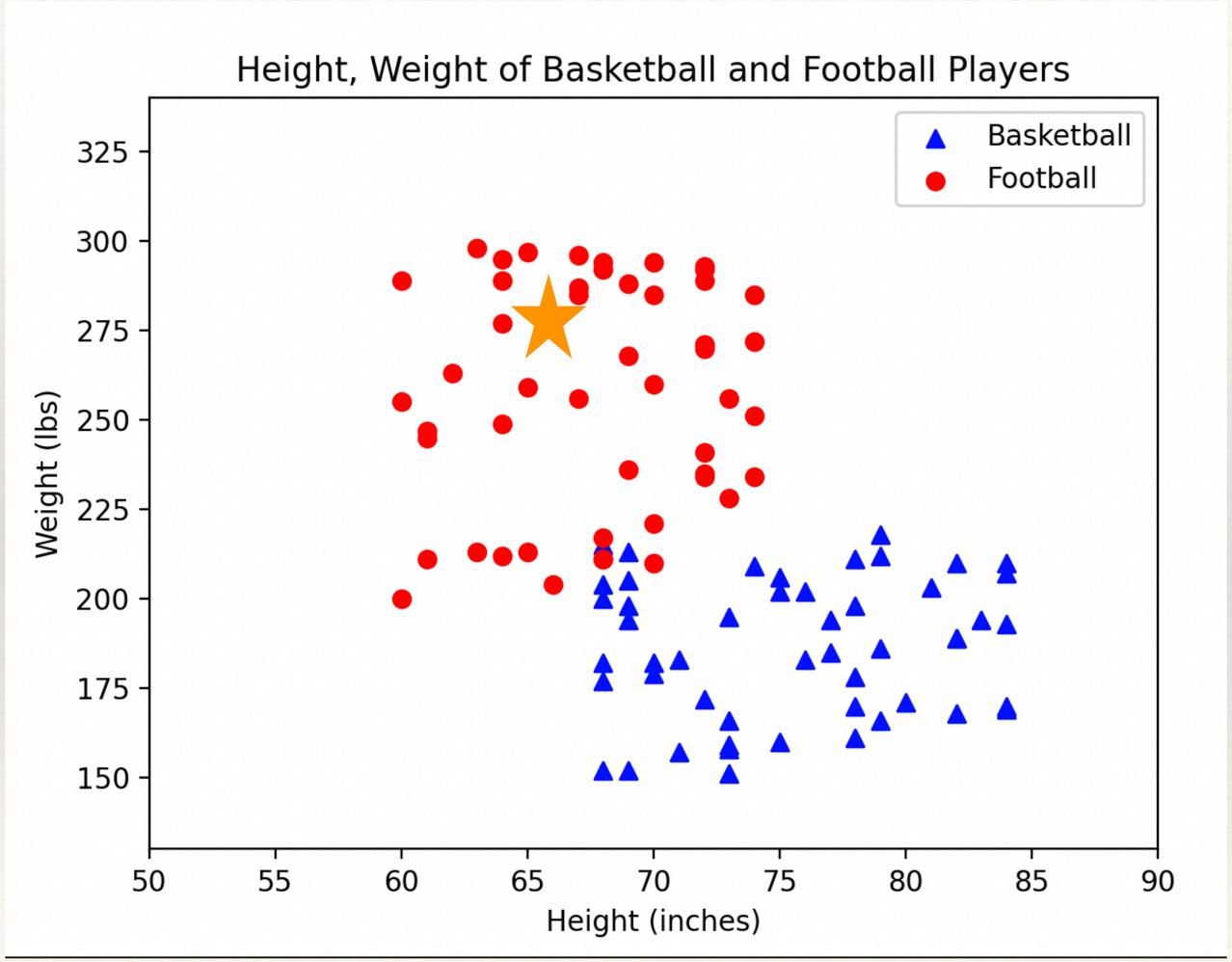
- Assign a data point to one of a predefined set of categories or classes
- For example:
  - Determining if an email is spam or not spam
  - Categorizing an image into types (e.g., cat, dog, horse)
  - Determining if an athlete is a football or basketball player

Feature Vector



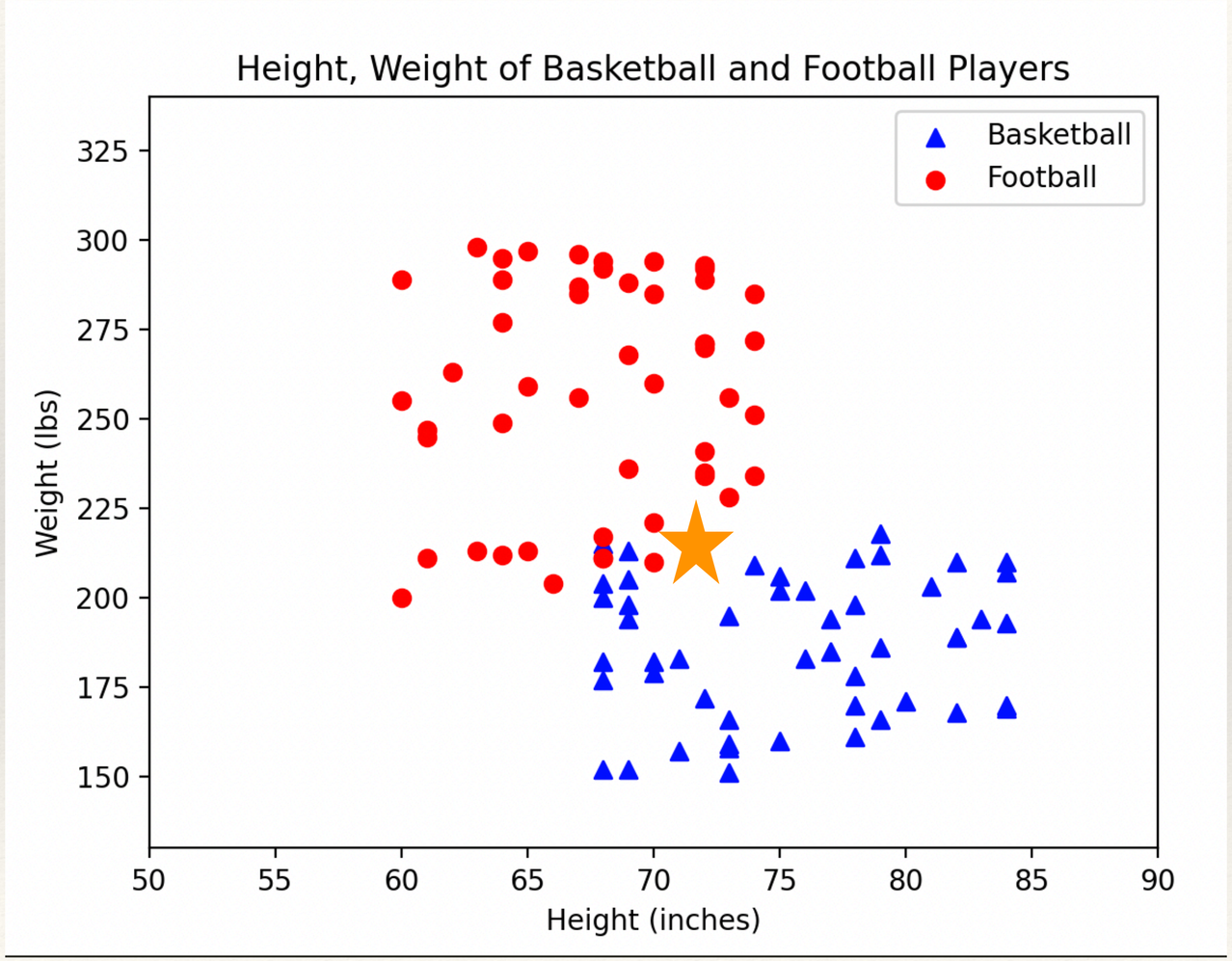
Classify the Orange star

Feature Vector



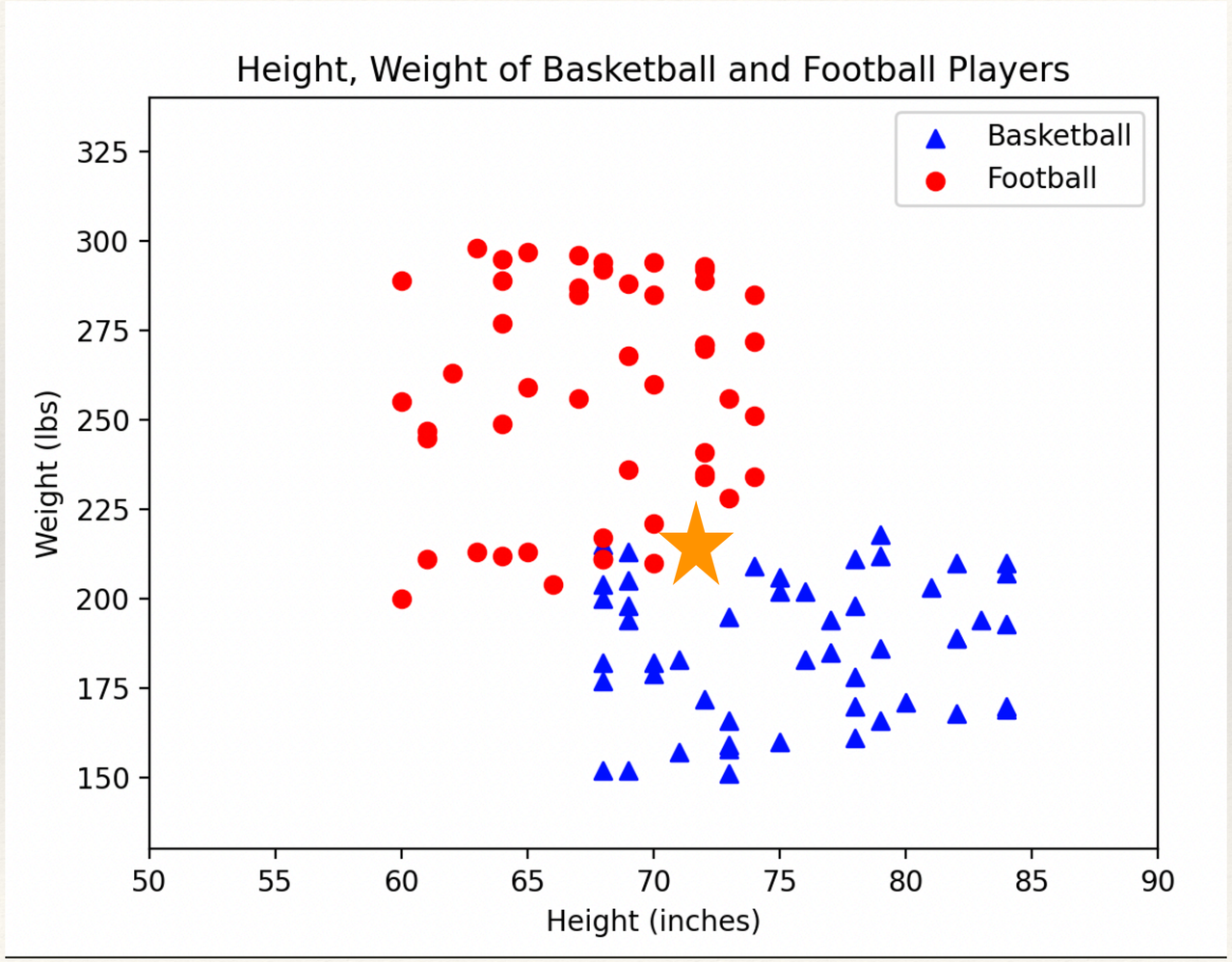
Classify the Orange star

Feature Vector



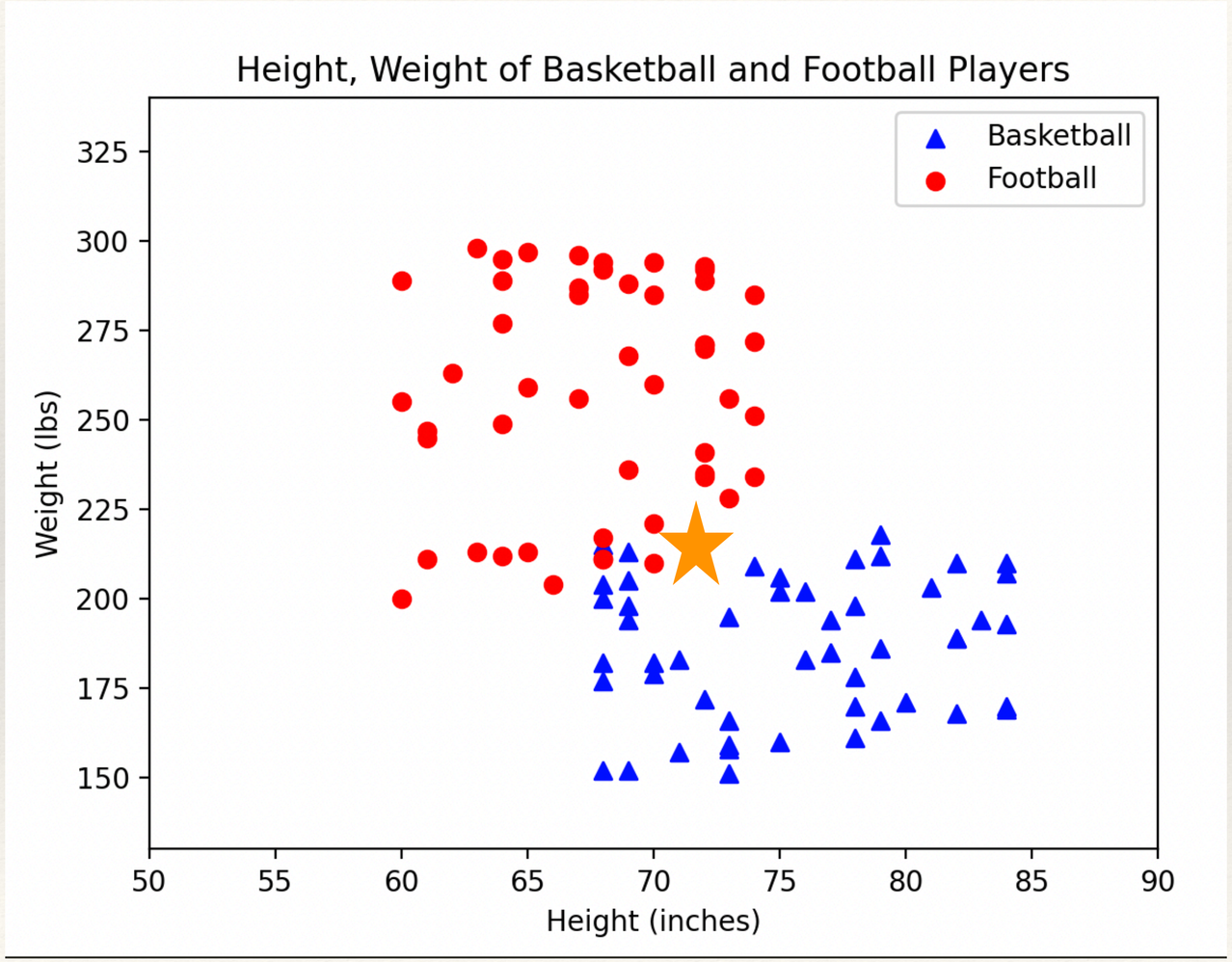
Classify the Orange star

K nearest neighbor (KNN)



KNN Algorithm
Find the K nearest neighbors
Predict the class the majority of those K neighbors belong to

K nearest neighbor (KNN)



Find the K nearest neighbors
Predict the class the majority of those K neighbors belong to

KNN Algorithm

KNN Algorithm

1. Choose the value of K
2. For a given point, calculate the distance to all of the training points
3. Identify the K nearest neighbors (those with the smallest distance to the point
4. The class that appears most frequently among the K neighbors is assigned to the point

KNN Algorithm

KNN Algorithm

1. Choose the value of K
2. For a given point, calculate the distance to all of the training points
3. Identify the K nearest neighbors (those with the smallest distance to the point
4. The class that appears most frequently among the K neighbors is assigned to the point

KNN Algorithm

KNN Algorithm

1. Choose the value of K
2. For a given point, calculate the distance to all of the training points
3. Identify the K nearest neighbors (those with the smallest distance to the point
4. The class that appears most frequently among the K neighbors is assigned to the point

For example, euclidean distance

$$d(A, B) = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2}$$

KNN Algorithm

KNN Algorithm

1. Choose the value of K
2. For a given point, calculate the distance to all of the training points
3. Identify the K nearest neighbors (those with the smallest distance to the point
4. The class that appears most frequently among the K neighbors is assigned to the point

Sort the neighbors by distance, look at the top K classes

KNN Algorithm

KNN Algorithm

1. Choose the value of K
2. For a given point, calculate the distance to all of the training points
3. Identify the K nearest neighbors (those with the smallest distance to the point
4. The class that appears most frequently among the K neighbors is assigned to the point

Examine the classes of the top K neighbors, assign the most common one to the new data point