Daily Machine Learning Interview Questions

16. What is a Decision Tree Classifier?

A decision tree builds classification (or regression) models as a tree structure, with datasets broken up into ever smaller subsets while developing the decision tree, literally in a tree-like way with branches and nodes. Decision trees can handle both categorical and numerical data.

17. What is Pruning in Decision Trees, and How Is It Done?

Pruning is a technique in machine learning **that reduces the size of decision trees. It reduces the complexity of the final classifier, and hence improves predictive accuracy by the reduction of overfitting.**

Pruning can occur in:

- **Top-down fashion. It will traverse nodes and trim subtrees starting at the root.**

- **Bottom-up fashion. It will begin at the leaf nodes**

There is a popular pruning algorithm called reduced error pruning, in which:

- Starting at the leaves, each node is replaced with its most popular class.
- If the prediction accuracy is not affected, the change is kept.
- There is an advantage of simplicity and speed.

18. Briefly Explain Logistic Regression ?

Logistic regression is a classification algorithm used to predict a binary outcome for a given set of independent variables.

The output of logistic regression is either a 0 or 1 with a threshold value of generally 0.5. Any value above 0.5 is considered as 1, and any point below 0.5 is considered as 0

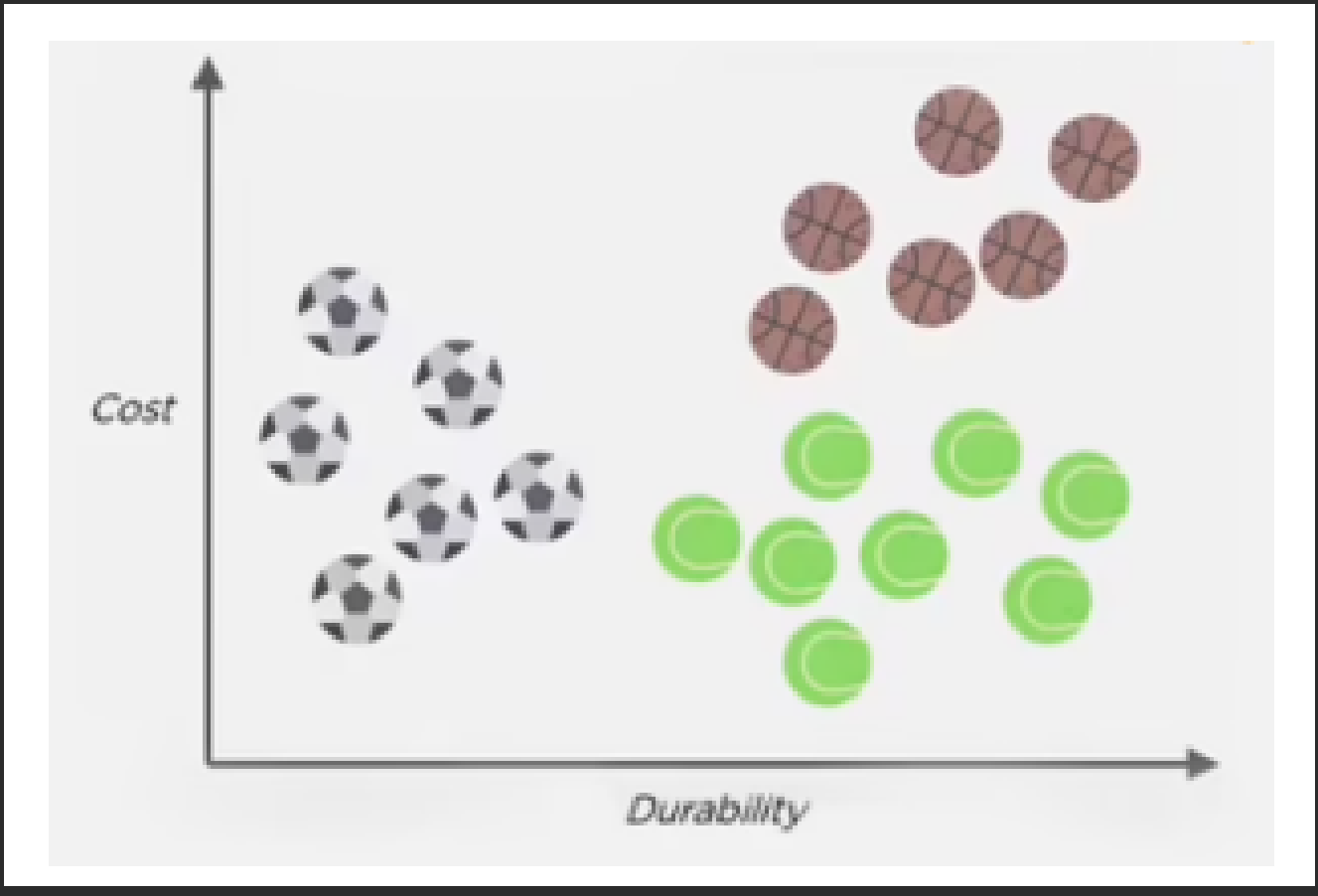
19. Explain the K Nearest Neighbor Algorithm.

K nearest neighbor algorithm **is a classification algorithm that works in a way that a new data point is assigned to a neighboring group to which it is most similar**.

In K nearest neighbors, K can be an integer greater than 1. So, for every new data point, we want to classify, we compute to which neighboring group it is closest.

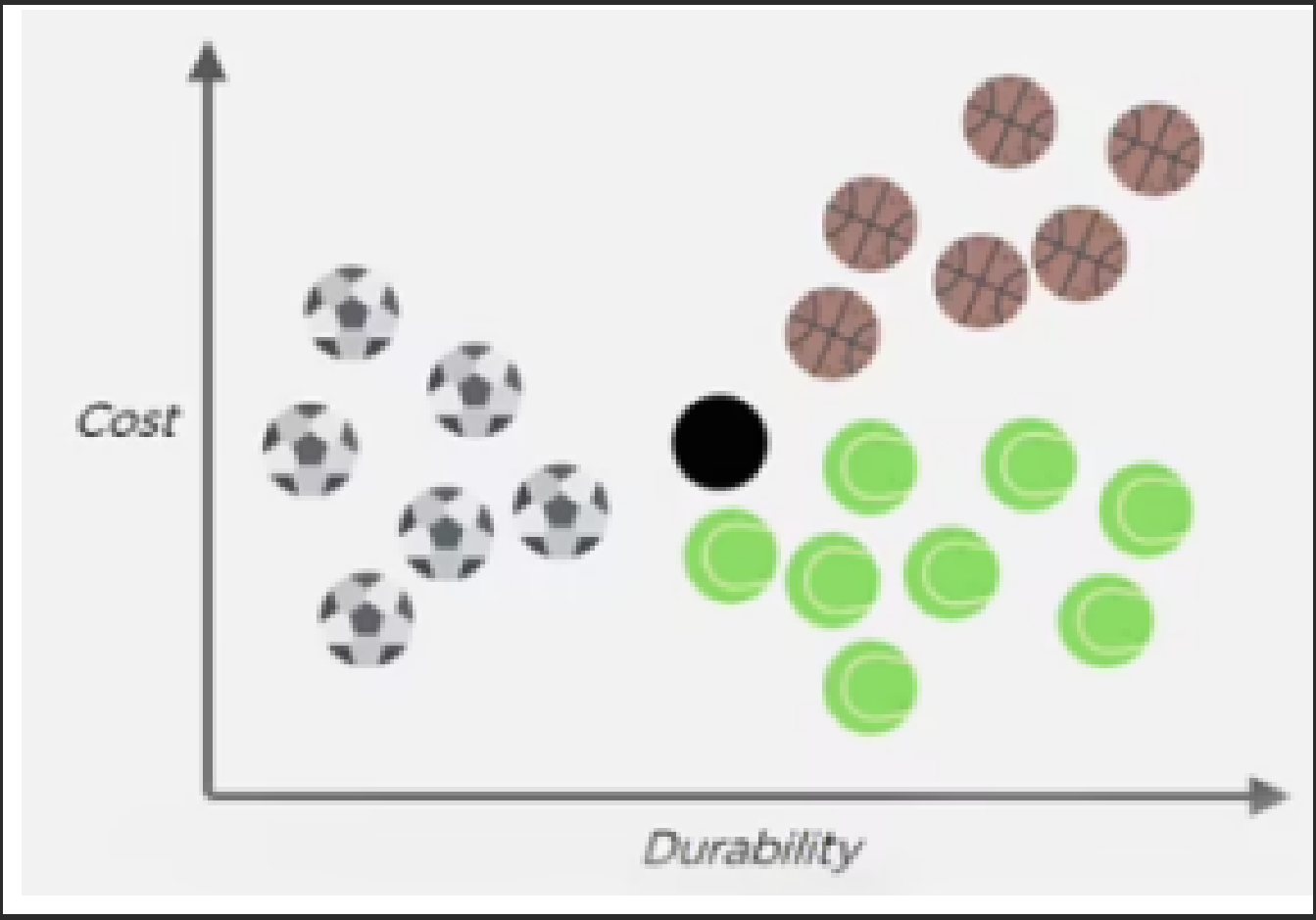
Let us classify an object using the following example. Consider there are three clusters:

- Football.
- Basketball.
- Tennis ball.



Let the new data point to be classified is a black ball. We use KNN to classify it. Assume K = 5 (initially).

Next, we find the K (five) nearest data points, as shown



Observe that all five selected points do not belong to the same cluster. There are three tennis balls and one each of basketball and football.

When multiple classes are involved, we prefer the majority. Here the majority is with the tennis ball, so the new data point is assigned to this cluster.

20. What is a Recommendation System?

Anyone who has used Spotify or shopped at Amazon will recognize a recommendation system:

It’s an information filtering system that predicts what a user might want to hear or see based on choice patterns provided by the user.

Thank You