Illustrative Example for Nearest Neighbour Search on a KD-Tree
Case of the College Athlete Dataset

Suppose we want to predict the target level of the query (6, 3.5) using 1-NN and the complete KD-tree shown in the lecture slides of Chapter 5.

Step 1: Follow the KD-Tree tests to "classify" the query.
This leads us to ID = 12 feature values (5, 2.5)
Calculate the Euclidean distance between ID12 and the query and consider it as the best distance and ID12 as the best instance:
bestDistance = distance = $\sqrt{(6-5)^2 + (3.5-2.5)^2} = 1.414$
bestInstance = ID12

Step 2: Go 1 level up in the tree. Found instance ID15 (4.75, 6.25)
distance = $\sqrt{(6-4.75)^2 + (3.5-6.25)^2} = 3.02$
Distance not better than *bestDistance*
Check if there could be other instances on the region of the split.
distanceToSplit = $|3.5 - 6.25| = 2.75$ Not smaller than *bestDistance* (i.e. the circle does not overlap with region beyond the split).
Decision: No descent on other branch.

Step 3: Go 1 level up in the tree. Found instance ID21 (6.75, 3)
distance = $\sqrt{(6-6.75)^2 + (3.5-3)^2} = 0.901$ < *bestDistance*
So *bestDistance* = 0.901
bestInstance = ID21
Check if there could be other instances on the region of the split.
distanceToSplit = $|6.75 - 6| = 0.75$ < *bestDistance* (i.e. the circle goes beyond the split to the other region). Need to explore…
Decision: descend using the other branch.

Step 4: Found instance ID20 (7.25, 5.75)
distance = $\sqrt{(6-7.25)^2 + (3.5-5.75)^2} = 2.5739$
Distance not better than *bestDistance.*
Check if there could be other instances on the region of the split.
distanceToSplit = $|5.75 - 3.5| = 2.25$ > *bestDistance.*
Decision: No descent on other branch.

Step 5: Go 1 level up in the tree. Found instance ID16 (5.5, 6.75)
distance = $\sqrt{(6-5.5)^2 + (3.5-6.75)^2} = 3.2882$
Distance not better than *bestDistance.*
Check if there could be other instances on the region of the split.
distanceToSplit = $|6.75 - 3.5| = 3.25$ > *bestDistance.*
Decision: No descent on other branch.

Step 6: Go 1 level up in the tree. Found instance ID6 (4.5, 5)
distance = $\sqrt{(6-4.5)^2 + (3.5-5)^2} = 2.1213$
Distance not better than *bestDistance.*
Check if there could be other instances on the region of the split.
distanceToSplit = $|6 - 4.5| = 1.5$ > *bestDistance*.
Decision: No descent on other branch.

Conclusion: The nearest neighbour is ID21.