

***k*-Nearest Neighbor**

Problem: The following table gives the Speed and Agility rankings of 10 football players from last year's graduating class, and whether they were drafted or not:

Players	Speed (1-10)	Agility (1-10)	Drafted?
1	2.5	6.0	No
2	2.0	5.5	No
3	3.0	7.5	No
4	3.5	5.0	No
5	4.0	4.0	No
6	2.0	2.0	No
7	8.0	8.5	No
8	5.0	6.0	Yes
9	5.0	9.5	Yes
10	7.5	8.0	Yes

The goal is to predict whether a new player will be drafted, using the *k*-Nearest Neighbor classification algorithm.

k-Nearest Neighbor algorithm:

1. Choose *k* and distance metric *d*
2. Find *k* nearest neighbors to sample
3. Assign class label via majority vote

Euclidean Distance metric

Similarity (distance) between instances x_i and x_j .
There are n attributes ($a_1..a_n$).

$$d(x_i, x_j) \equiv \sqrt{\sum_{z=1}^n (a_z(x_i) - a_z(x_j))^2}$$

1. The new player has rankings Speed=7.0 and Agility=3.0. Determine the distance d between the new player and last year's players:

Example: distance between new player and Player 1:

$$d = \sqrt{(7.0 - 2.5)^2 + (3.0 - 6.0)^2} = 5.41$$

Players	Distance
1	5.41
2	5.59
3	6.02
4	4.03
5	3.16
6	5.10
7	5.59
8	3.61
9	6.80
10	5.02

2. For $k=1$, how is the new player classified:

Nearest neighbor is Player 5 (3.16) \Rightarrow classification = No.

3. For $k=3$, how is the new player classified:

Nearest 3 neighbors are:

Player 5 (3.16) No

Player 8 (3.61) Yes

Player 4 (4.03) No

Use majority vote to assign label:

\Rightarrow classification = No.