# “Unpeeling” k-Nearest Neighbors



## **The *k*-Nearest Neighbors Algorithm**

One method of classifying the *Goldy Fruit* is to use an algorithm called *K-*Nearest Neighbors (kNN). The main idea is that an unclassified case is classified based on the *k* cases that are most similar to it using majority "vote". The algorithm for this is:

**Step-1:** Select the number *k* of the neighbors

**Step-2:** Calculate a similarity measure between the unclassified case and all the classified cases

**Step-3:** Identify the *k* nearest neighbors as per the calculated similarity measure.

**Step-4:** Among these *k* neighbors, count the number of cases in each class (i.e., category).

**Step-5:** Assign the unclassified case to the majority class within the *k* nearest neighbors (the one with the most cases)

1. Use the kNN algorithm and *k* = 3, to determine what type of fruit the *Goldy Fruit* may be. Also report the “vote” tally for each of the four classes (i.e., fruit types).

| **Name** | **Type** | **Number of Seeds** | **Similarity** |
| --- | --- | --- | --- |
| Blackberry | Berry | 94 | 84 |
| Blueberry | Berry | 50 | 40 |
| Raspberry | Berry | 110 | 100 |
| Grapefruit | Citrus | 40 | 30 |
| Lemon | Citrus | 8 | -2 |
| Orange | Citrus | 10 | 0 |
| Apricot | Drupe | 1 | -9 |
| Cherry | Drupe | 1 | -9 |
| Peach | Drupe | 1 | -9 |
| Apple | Pome | 6.5 | -3.5 |
| Asian pear | Pome | 5 | -5 |
| Pear | Pome | 10 | 0 |

Pear = 0

Orange = 0

Lemon = -2

Vote Tally:

Citrus = 2

Pome = 1

Based on this k value we would classify it as a citrus.

**STOP: Raise your hand to check in with us.**

1. Use the kNN algorithm and *k* = 5, to determine what type of fruit the *Goldy Fruit* may be. Also report the “vote” tally for each of the four classes (i.e., fruit types).

Pear = 0

Orange = 0

Lemon = -2

Apple = -3.5

Asian Pear = -5

Vote Tally:

Citrus = 2

Pome = 3

Based on this k value we would classify it as a pome.

1. Use the kNN algorithm and *k* = 7, to determine what type of fruit the *Goldy Fruit* may be. Also report the “vote” tally for each of the four classes (i.e., fruit types).

Pear = 0

Orange = 0

Lemon = -2

Apple = -3.5

Asian Pear = -5

Apricot = 9

Cherry = 9

Vote Tally:

Citrus = 2

Pome = 3

Drupe = 2

This k value will have us classify it as a pome.

## 

## 

## 

## 

## 

## **Part IV: Choosing the Value of *k***

In the previous section, you saw that different values of *k* can result in different classifications for the *Goldy Fruit.* In practice we would have to evaluate which *k* value is optimal by testing some cases for which we know the correct classification. We would then choose the *k* value that seemed to consistently classify these cases correctly. Rather than have you do that, we are going to determine the *k* value via the following formula, which works well in most situations.



where *n* is the number of cases for which we know the classification, and  is the ceiling function (round up to the nearest integer).

1. Compute the value for *k* for our fruit classification using the formula. Show your work.

sqrt(12) = 3.46 = 4

Optimal k = 4

1. Use the optimal number of nearest neighbors (that you reported in the previous question) to determine what type of fruit the *Goldy Fruit* may be. Also report the “vote” tally for each of the four classes (i.e., fruit types).

Pear = 0

Orange = 0

Lemon = -2

Apple = -3.5

Vote Tally:

Citrus = 2

Pome = 2

It may be a citrus or a pome based on this classification.