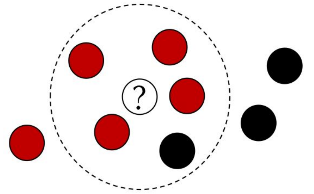
* Establishing Shot – Ranger in a National Park with BooBoo. They’re examining a flower but don’t know what species it is…
* Ranger calls Acme Consulting to ask for help. StatCat & DataDog arrive at the park and the Ranger shares the classification problem with them.
* Von Drake Lecture

StatCat & DataDog confer and then StatCat comes back and sets up a portable lectern. He first tells Yogi how he and DataDog used the K-Nearest Neighbors algorithm to classify Iris flowers. He then lectures on the basic ideas behind this algorithm, using the information in blue below.

K-Nearest Neighbors (k-NN) is an algorithm that classifies a data point based on the classification of its neighbors. This means that if a data point is surrounded by four red points and one black point, then majority vote would suggest that the data point is likely red.



The k in k-NN is a parameter referring to the number of nearest neighbors to include in the majority vote. In the example above, k equals five and includes all of the colored circles within the dashed line. Given this scenario, what color do you think the algorithm will predict for the circle with the question mark? Yes, the algorithm will predict that it’s red.

But how many neighbors should be considered? That is, what value should k be set to? Choosing the right value of k is a process known as parameter tuning, and it is critical to prediction accuracy.

The Ranger is duly impressed. Indeed, he believes that the unknown flower in the park is a relative of the Iris and is certain that the algorithm will work in this case too.

* Data Collection Scene

StatCat, the Ranger, and BooBoo head out to collect samples of the park’s flowers and return with a basket of what appear to be three distinct species. Meanwhile, DataDog builds a miniature A.I. frog to measure each of the flowers and then generate an Iris-like dataset.

* Data Learning Event
* A.I. frog takes measurements of the unknown species, feeds it to the algorithm, and it responds with a prediction. Everyone is excited to finally know what species this flower is. But while focused on the computer output, the unknown flower swallows DataDog’s A.I. frog. Apparently, this new species is a carnivore, a member of the Venus Flytrap family!

A.I. Tour of Applications

Flower Story – K-Nearest Neighbors

Goldilocks Story – Overfit, Underfit, and Just Right