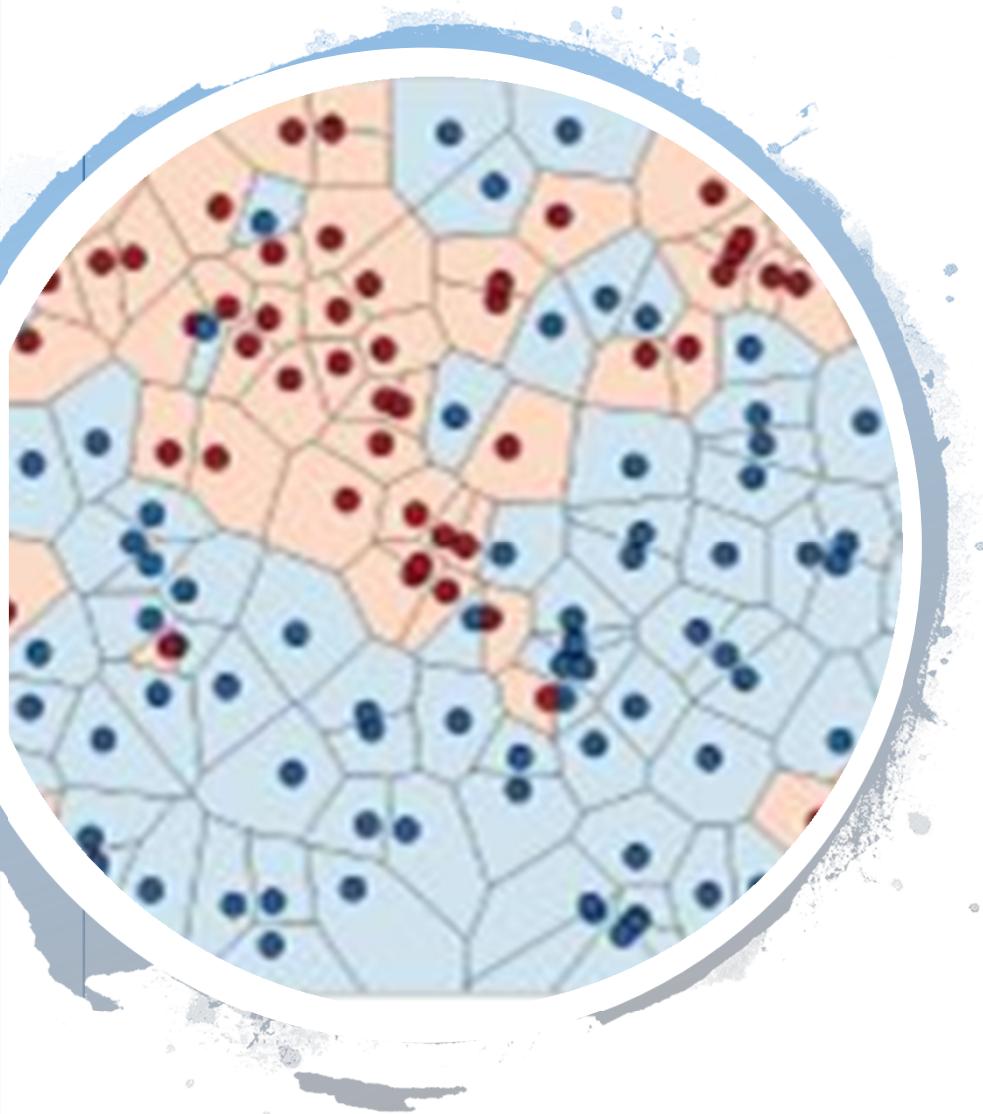


K Nearest Neighbor



What is KNN?

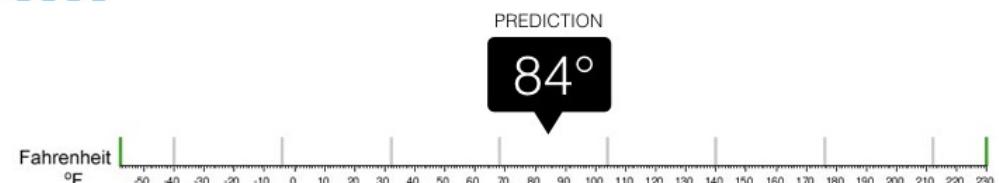
- It is one of the simplest **Supervised Machine Learning** algorithm.
- K nearest neighbors **stores all available cases** .
- It classifies new cases based on a **similarity measure**(eg; distance).

- It is used for both **classification and regression** predictive problems.
- However, it is more widely used in **classification problems** in the industry.



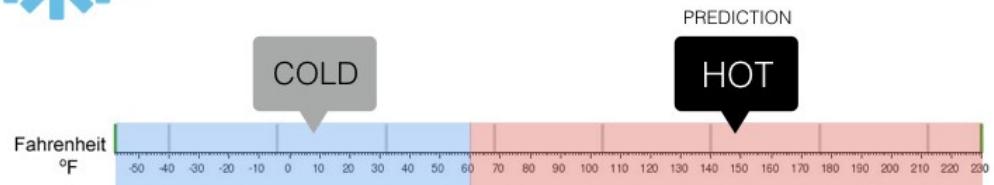
Regression

What is the temperature going to be tomorrow?



Classification

Will it be Cold or Hot tomorrow?



KNN is based
on Feature
similarity



Lets consider an example:





CATS



Sharp Claws, uses to climb

Smaller length of ears

Meows and purrs

Doesn't love to play around

DOGS

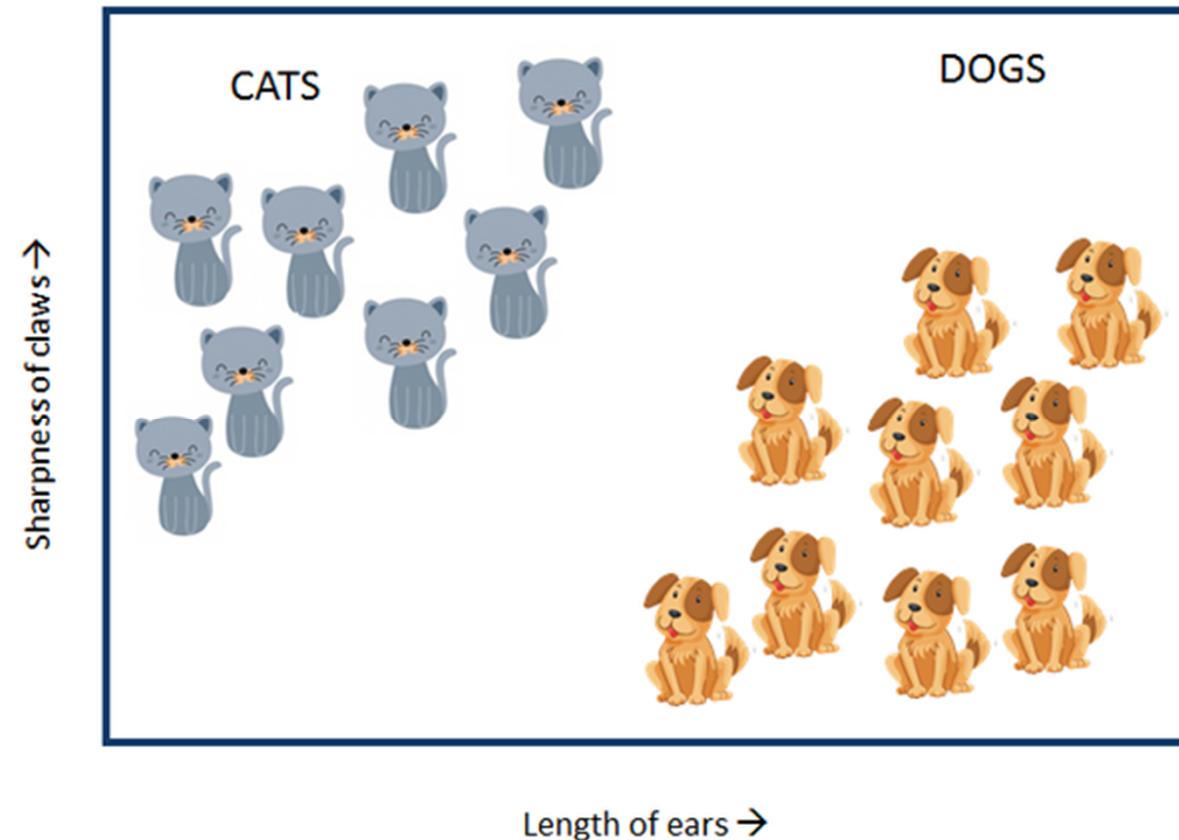


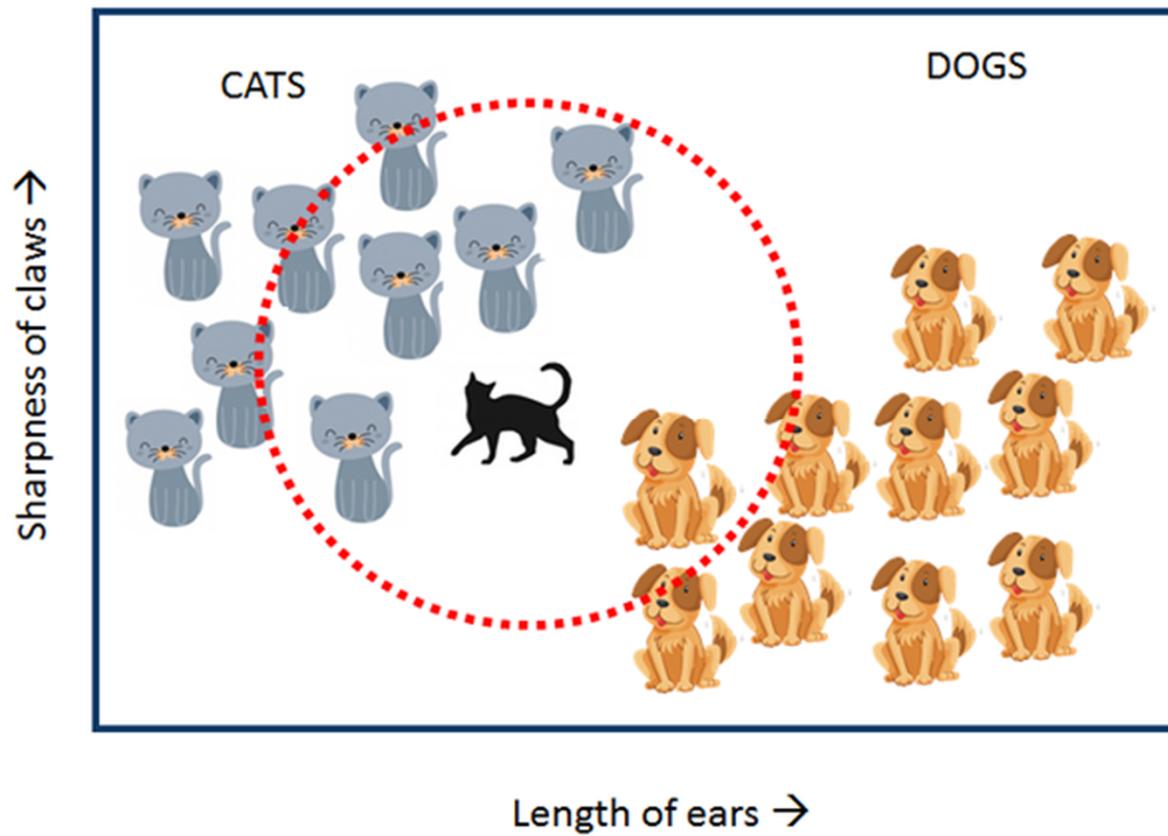
Dull Claws

Bigger length of ears

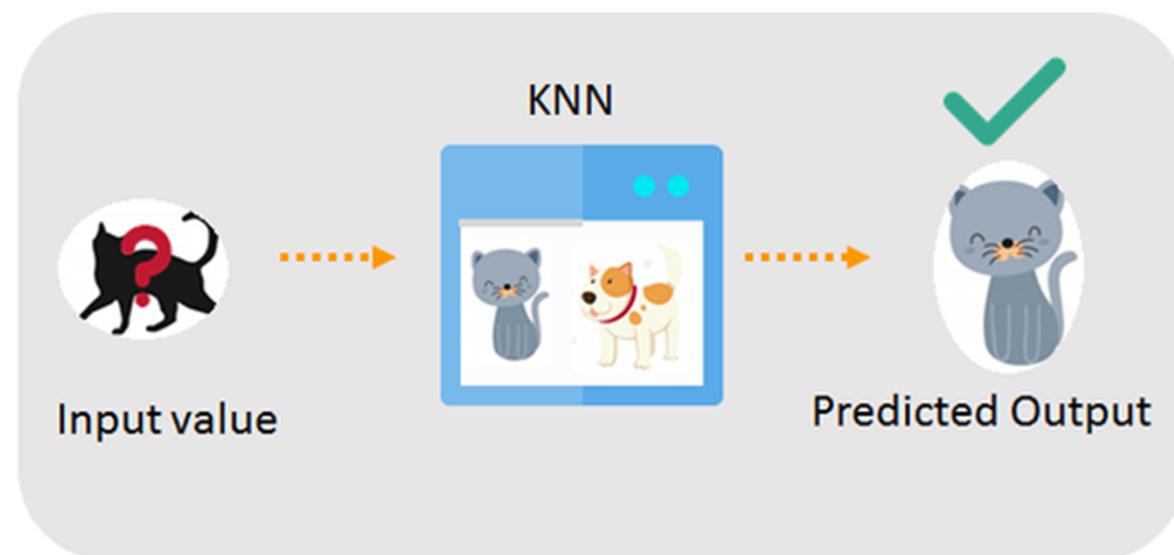
Barks

Loves to run around

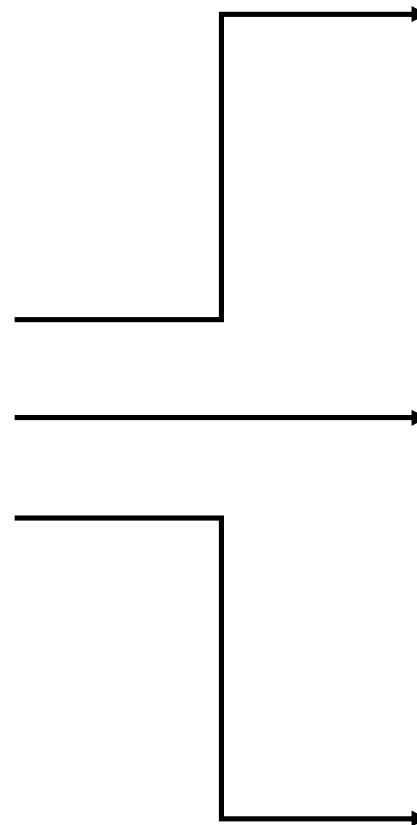




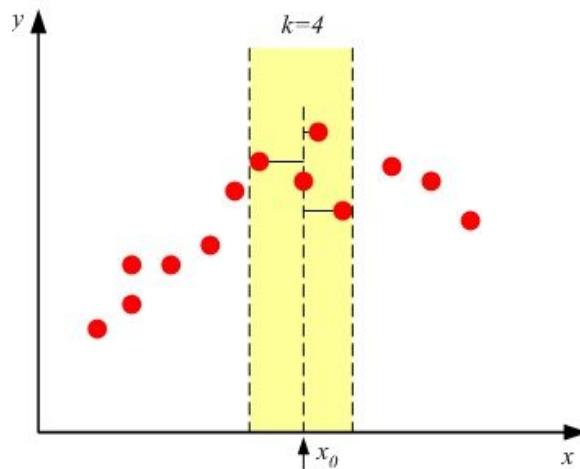
- Therefore, **the features** of the strange animal are **more like cats**.
- K nearest neighbors is a simple algorithm that **stores all available cases** (cats and dogs) and **classifies new cases** (unknown animal) based on a similarity measure or characteristics.
- Hence, it is proved that the strange animal is a cat.



KNN is also called

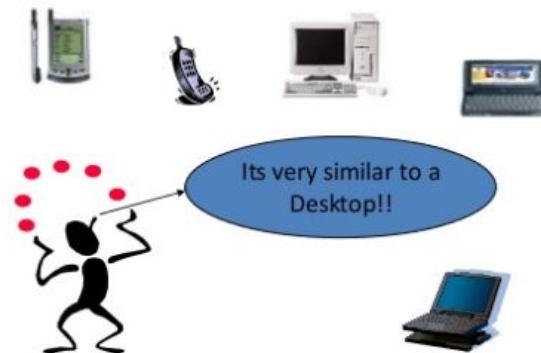


Non - parametric Learning



- A non parametric model is one that can not be characterized by a **fixed set of parameters**.
- It means that it **does not make any assumptions** on the **underlying data distribution**.

Instance Based Learning



- It is based on the **memorization** of the dataset.
- The **number of parameters** is **unbounded** and **grows with the size** of the data.
- It **classifies new instances** based on direct comparison.
- It is also referred as **Lazy Learning** because they **delay processing** until a new instance is classified

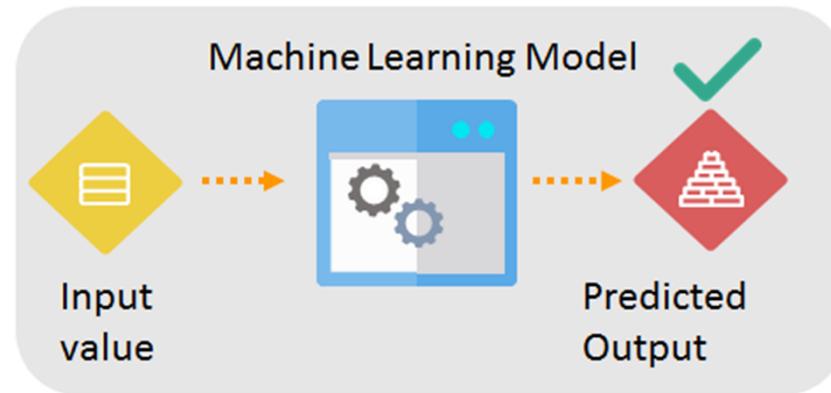
Lazy Learning



- Lack of generalization because KNN **keeps all the training data.**
- In other words, there is **no explicit training phase**

Why KNN?

- By now, we all know Machine learning models **makes predictions** by learning from the **past data available**.





LIKE A GOOD NEIGHBOR



STAY OVER THERE

More pics on: www.imfunny.net

K Nearest Neighbor Algorithm

- **Step 1:** Choose a value for K. K should be an odd number.
- **Step 2:** Find the **distance of the new point** to each of the training data.
- **Step 3:** Find the **K nearest neighbors** to the new data point.
- **Step 4:**
 - **For classification**, count the number of data points in each category among the k neighbors. **New data point** will belong to class that has **the most(Mode) neighbors**.
 - **For regression**, value for the new data point will be the **average(Mean) of the k neighbors**.

What is K is K nearest neighbors?

- K is a number used to **identify similar neighbors** for the new data point.
- KNN takes K nearest neighbors to decide where the **new data point** with belong to.
- This decision is based on **feature similarity**.

For example

- We have Friend circle in **our new neighborhood**.
- We select 3 neighbors that we want to be **close friends based on common thinking or hobbies**.
- In this case **K is 3**.

