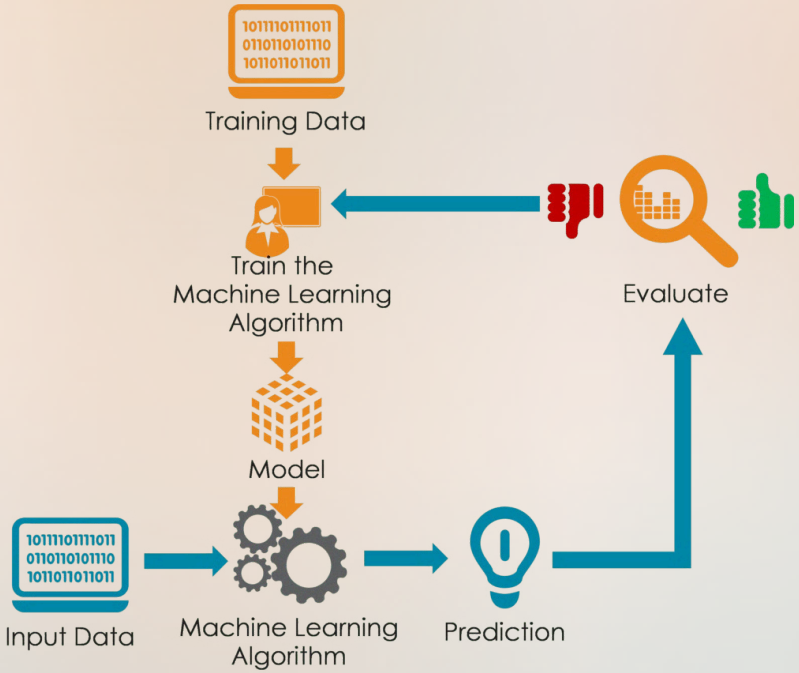


Introduction to KNN Classification

K-Nearest Neighbors (KNN) is a powerful and versatile algorithm in machine learning. It's a non-parametric method used for classification and regression tasks.

Made with Gamma

What is KNN?

KNN is a supervised learning algorithm that classifies data points based on the majority class of their nearest neighbors. It assumes that similar data points are likely to belong to the same class.

Non-Parametric

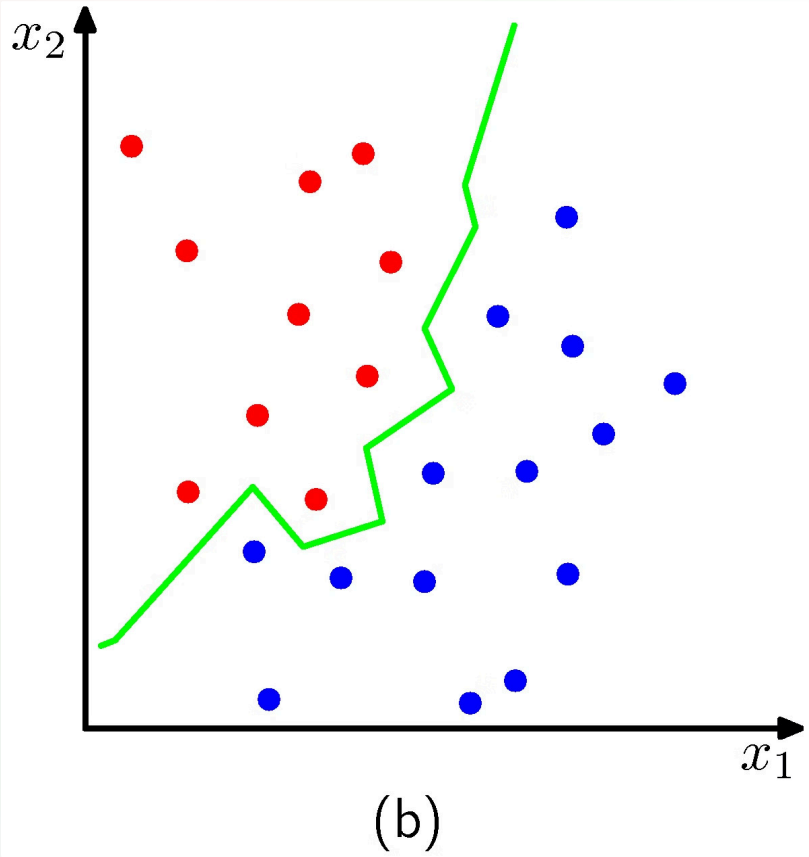
KNN doesn't make assumptions about the underlying data distribution.

Lazy Learning

KNN does not build a model explicitly, but instead relies on the training data during prediction.

Instance-Based

It makes predictions based on the similarity to known data points, rather than a model.



Made with Gamma

How does KNN work?

KNN works by finding the k nearest neighbors to a new data point and assigning it to the class that is most frequent among those neighbors.

1

Calculate Distance

Measure the distance between the new data point and all existing data points in the training set.

2

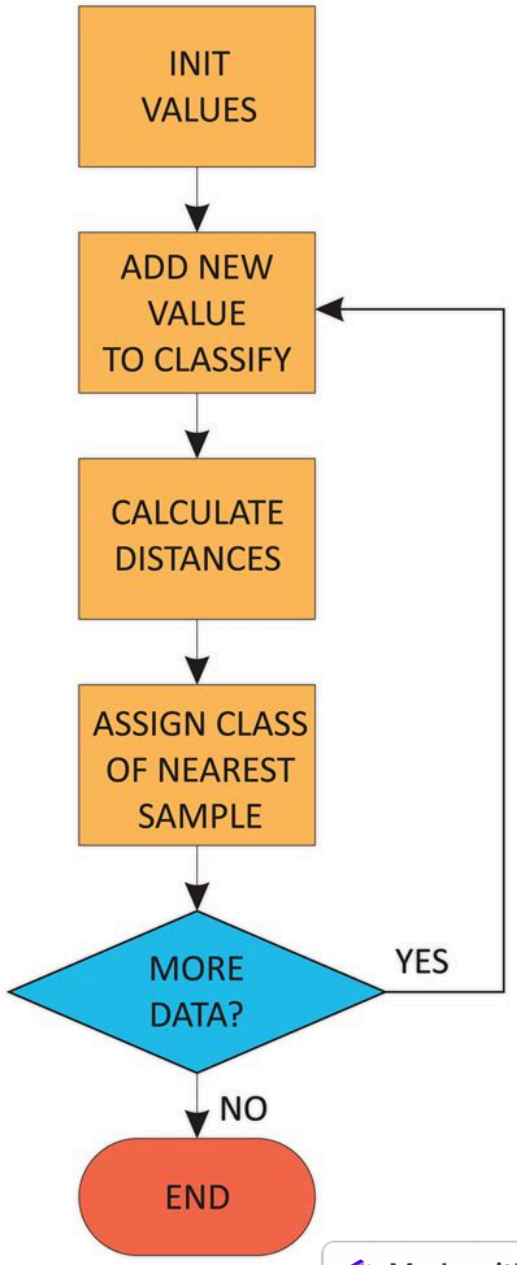
Find Nearest Neighbors

Identify the k data points closest to the new data point based on the calculated distances.

3

Assign Class

Determine the most common class among the k nearest neighbors and assign it to the new data point.



Made with Gamma

Choosing the right value of K

The value of K, the number of nearest neighbors to consider, is a crucial hyperparameter that affects KNN's performance.

1

Small K

More prone to noise and overfitting, as it focuses on local patterns.

2

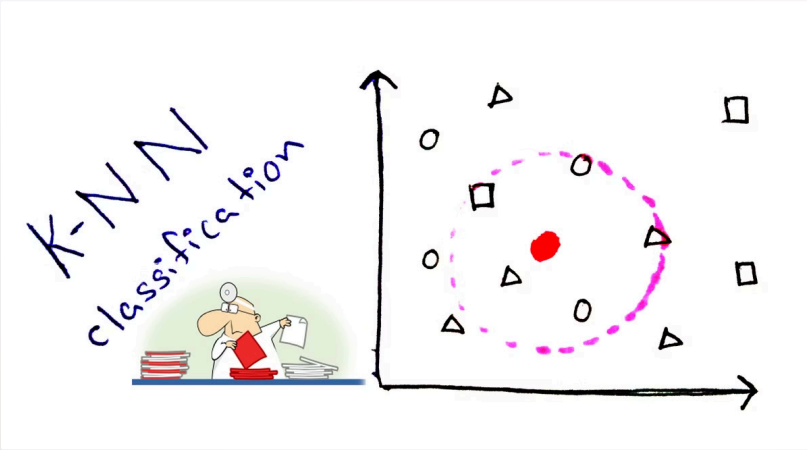
Large K

More robust to outliers but can lead to underfitting, as it considers a wider range of data points.

3

Optimal K

A balance between bias and variance, resulting in a model that generalizes well to new data.



Made with Gamma

Algorithm Steps

Distance Metric: The common distance metrics used based on Nature of the Data.(Euclidean distance,Manhattan distance)

1 **Choose the number of neighbors (k).**

2 **Calculate the distance between the query point and all other points.**

3 **Sort the distances and determine the nearest neighbors.**

4 **Assign the most common class among the nearest neighbors to the query point.**

Made with Gamma

Advantages of KNN

KNN offers several advantages that make it a popular choice for various machine learning applications.

1 Simplicity

Easy to understand and implement. Also no training phase is reuired.

2 Versatility

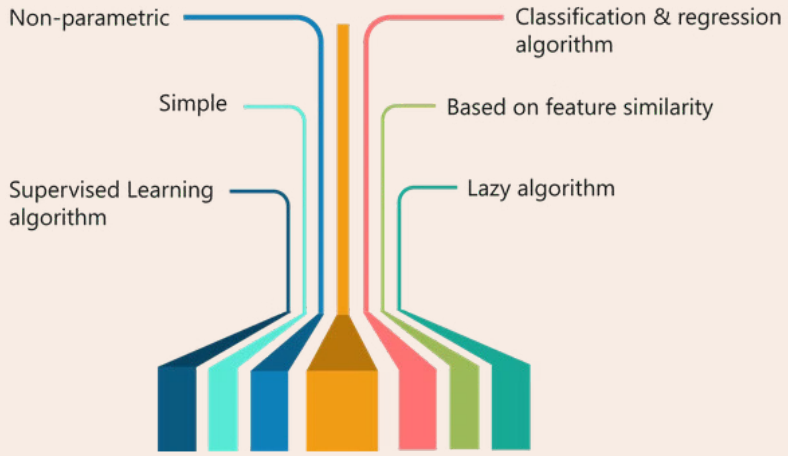
It can be used for both classification and regression tasks.

3 Non-Parametric

It doesn't assume a specific data distribution, making it suitable for complex datasets.

4 Easy to Update

New data points can be easily added to the training set without retraining the entire model.



Made with Gamma

Disadvantages of KNN

Computational Complexity

It requires calculating distances to all training data points during prediction, which can be computationally expensive for large datasets.

Sensitive to Noise

Outliers or noisy data points can significantly influence the classification result, affecting accuracy.

Curse of Dimensionality

In high-dimensional datasets, distances become less meaningful, potentially leading to inaccurate results.

Made with Gamma

Implementing KNN in Python

Implementing KNN in Python is relatively simple using libraries like scikit-learn.

Import Libraries

Import necessary libraries for data manipulation, preprocessing, and KNN implementation.

Load Data

Load your dataset into a suitable data structure for processing and model training.

Preprocess Data

Clean and transform your data to ensure it's suitable for KNN, including scaling, imputation, and outlier handling.

Train Model

Initialize a KNN model with your chosen parameters and train it using the preprocessed data.

Made with Gamma

Applications of KNN

KNN has numerous applications in various domains, demonstrating its versatility and practical value.

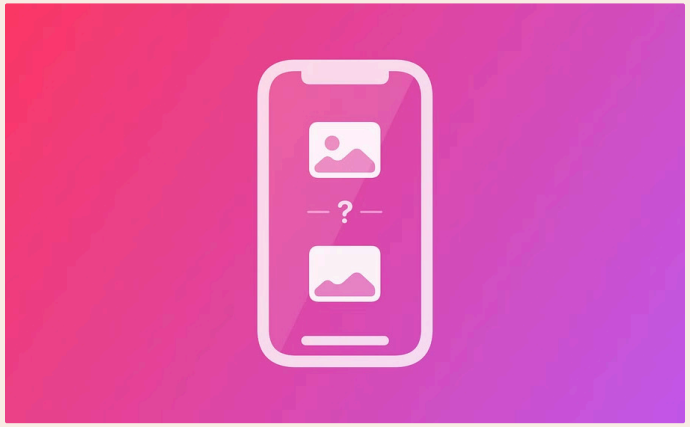
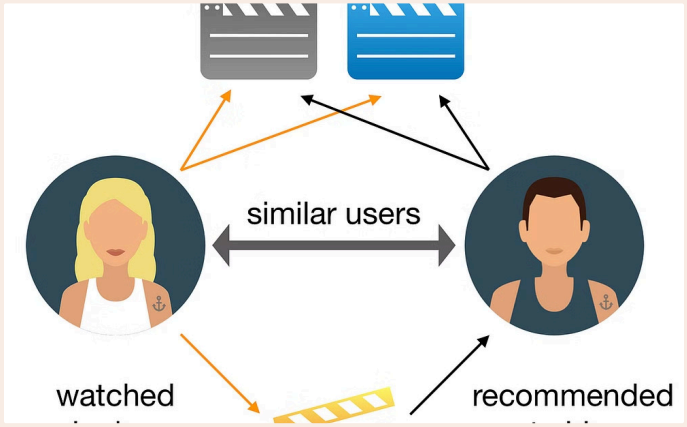


Image Recognition

Classifying images based on visual features, such as object identification or scene analysis.



Recommender Systems

Predicting user preferences for products or services based on similar items or users.

Medical Diagnostics

Identifying diseases or conditions based on patient data, such as symptoms or medical history.

Made with Gamma