eCornell

CHEAT SHEET

k-Nearest Neighbors

Algorithm Name	k-NN
Description	For a given test point, the k-NN algorithm identifies the k most similar training points and finds the most common label amongst them. This label is used as a prediction for the test point.
Applicability	Often competitive in low-dimensional spaces in settings with many classes. Used for classification or regression.
Assumptions	"Similar inputs have similar labels"; k-NN assumes that the user has a way to compute distances that reflect meaningful dissimilarities.
Underlying Mathematical Principles	• Distance metrics
Additional Details	• Hyperparameter is number of neighbors (k) • Dealing with ties - fall back to smaller k values • Special data structures such as k-d tree or ball tree can make k-NN more efficient • Distance metric used is application-specific • For image classification, we can rotate, translate the images to extend the training set
Example	Identify individuals visible in a photo uploaded to a social media account.