

En Yakın k-Komsu (k-Nearest Neighbor)

Yapay Ogrenim alanında örnek bazlı öğrenen algoritmalarından bilinen kNN, eğitim verinin kendisini sınıflama (classification) amaçlı olarak kullanır, yeni bir model ortaya çıkartmaz. Bu amaçla etiketleri bilinen eğitim verisi bir kenarda tutulur ve yeni bir veri noktası alınıncaya bu veriye geri donulur ve o noktaya “en yakın” k tane nokta bulunur. Daha sonra bu noktaların etiketlerine bakılır ve genellikle çoğunluğun etiketi ne ise, o etiket yeni noktanın etiketi olarak kabul edilir.

“En yakın” sözü bir koordinat sistemi çağrıştırır ve evet, kNN, aynen k-Means ve diğer pek çok sayısal bazlı öğrenme yöntemi gibi, eldeki çok boyutlu veri noktalarının elemanlarını bir koordinat sistemindeymiş gibi kabul eder. Kiyasla mesela APriori gibi bir algoritma metin bazlı veriyle olduğu gibi çalışabilir.

Bir veri obgesi içinden en yakın noktaları bulmanın en basit yolu bu listeyi baştan sonra taramak (kaba kuvvet yöntemi -brute force-) ve her listedeki nokta ile yeni nokta arasındaki mesafeyi teker teker hesaplayıp en yakın k taneyi içinden seçmektir. Fakat bu işlem de hızlandırılabilir.

