

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. Algoritma şöyle işler: etiketleri bilinen eğitim verisi alınır ve bir kenarda tutulur. Yeni bir veri noktası görüldüğünde bu veriye geri dönülür ve o noktaya “en yakın” k tane nokta bulunur. Daha sonra bu noktaların etiketlerine bakılır ve çoğunluğun etiketi ne ise, o etiket yeni noktanın etiketi olarak kabul edilir.

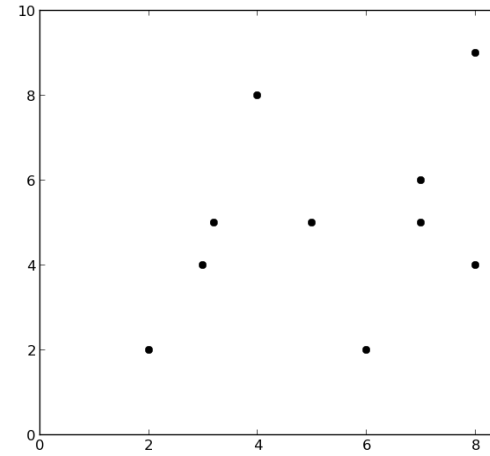
“En yakın” sözü bir koordinat sistemi anlamına geliyor, ve kNN, aynen k-Means ve diğer pek çok koordinatsal öğrenme yöntemi gibi eldeki çok boyutlu veri noktalarının elemanlarını bir koordinat sistemindeymiş gibi görür. Kiyasla mesela APriori gibi bir algoritma metin bazlı veriyle olduğu gibi çalışabilirdi.

Peki arama bağlamında, bir veri obgesi içinden en yakın noktaları bulmanın en basit yolu nedir? Listeyi bastan sonra taramak (kaba kuvvet yöntemi -brute force-) ve her listedeki nokta ile yeni nokta arasındaki mesafeyi teker teker hesaplayıp en yakın k taneyi içinden seçmek bir yöntem. Bu basit algoritmanın yuku $O(N)$ 'dir.

Fakat bu işlemi daha hızlı yapmak mümkün.

Arama algoritmaları kullanarak eğitim verilerini bir ağacı yapısı üzerinden arayarak erişim hızını $O(\log N)$ 'e indirmek mümkündür.

Küresel Ağaçlar (Ball Tree)



Diyelim ki elimizde iki boyutlu uzayda alttaki gibi bir veri var.