KÜMELEME

Birbirlerine benzeyen veri parçalarını ayırma gruplama işlemidir. Temel yaklaşım Öklid Manhattan Minkowski uzaklık bağıntıları kümeleme işleminde alt işlem olarak kullanılır.

K en Yakın komşu algoritması ve en uzak komşu bilinen yöntemlerdir. Hiyerarşik

K means yöntemi Hiyerarşik olmayan

Uzaklık ölçütleri

Kümeleme yöntemlerinin çoğu gözlem değerleri arasında ki uzaklıkların hesaplanması esasına dayanır. Çeşitli ü değişken 5 gözlemden oluşan X matrisi verilsin

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \\ x_{51} & x_{52} & x_{53} \end{bmatrix}$$

x_{11} x_{12} x_{13} Birinci gözlem

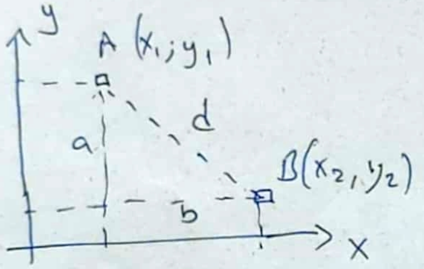
x_{21} x_{22} x_{23} İkinci " Bu iki gözlem arası uzaklık ise $d(2,1)$ biçiminde yazılabilir.

Yukarıdaki X matrisinin her bir satırının diğerine olan uzaklığı $d(i,j)$ biçiminde ifade edilecek olursa D uzaklıklar matrisi

$$D = \begin{vmatrix} 0 & & & & \\ d(2,1) & 0 & & \text{Simetrik} & \\ d(3,1) & d(3,2) & 0 & & \\ d(4,1) & d(4,2) & d(4,3) & 0 & \\ d(5,1) & d(5,2) & d(5,3) & d(5,4) & 0 \end{vmatrix}$$

Yukarıdaki matrisin üst kısmı alt kısmının simetriğidir. $d(i,j) = d(j,i)$ Birden fazla uzaklık bağıntısı vardır. Yaygın olan üçtanesi

a) Öklid



$$d(A,B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \qquad d(i,j) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}$$

b) Manhattan

$$d(i,j) = \sum_{k=1}^{p} \left(|x_{ik} - x_{jk}| \right) \qquad i,j = 1,2,\dots n \qquad k = 1,2,\dots p$$

c) Minkowski uzaklığı

$$d(i,j) = \left[\sum_{k=1}^{p} |x_{ik} - x_{jk}|^m \right]^{1/m} \qquad i,j = 1,2,\dots,n \qquad k = 1,2,\dots,p$$

$m=2$ için öklid