

Vektör Uzay Modeli

D1 ve D2 gibi iki doküman, dizinleme sonrasında artık dizin terimleri ile temsil edilir.

$$D1=(a, b, c, d, e, f) \quad D2=(a, c, z)$$

Bu dokümanlar birer nokta olarak düşünüldüğünde



Eğer bu iki doküman birbirlerine benziyorlarsa yani ortak terimleri fazlaysa aralarındaki mesafe kısa olacaktır.



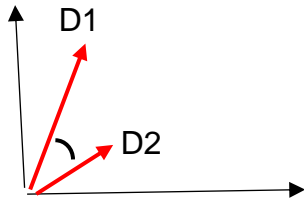
Eğer bu iki doküman birbirlerine benzemiyorlarsa, yani ortak terimleri ya hiç yok ya da az ise aralarındaki mesafe fazla olacaktır.

Vektör uzayında bu dokümanlar vektörler olarak temsil edildiğinde “mesafe” vektörlerin birbirlerine olan açıları dikkate alınarak bir benzerlik fonksiyonu olarak ifade edilebilir.

Bu benzerlik, karşılıklı olarak dokümanlardaki terimlerin benzerliğini ve terim ağırlıklarını yansıtır. Bu tür bir benzerlik ölçümü, iki vektörün içsel ürünü (terim benzerliği) veya alternatif olarak vektör çifti arasındaki açının bağıl fonksiyonu ile yapılabilir.

Her iki vektöre atanan terimler aynı olduğu zaman yani iki doküman tam bir benzerlik gösterdiğinde aralarındaki **açı 0** olacak ve en yüksek benzerliği üretecektir.

Aralarındaki açının esas alındığı ve koordinat sisteminde her vektörün 0 noktasından başlayarak tanımlandığı benzerlik ölçümü yerine, vektörler arasındaki bağıl / görel **mesafenin** bütün vektör uzunluklarının tek bir uzunluğa indirildiği (uzunluğun normalize edildiği) ve uzayın yansımaları biçiminde vektörlerin tek bir alanda temsil edileceği bir hesaplama dayalı bir yaklaşım dikkate alınabilir.



O zaman benzer dizin terimlerine sahip iki doküman, uzayda çok yakın noktalarda bulunacaktır. Genel olarak **iki doküman noktası arasındaki mesafe** ile **vektörler arasındaki benzerlik** birbiriyle **ters orantılıdır**. Yani aralarındaki mesafe azaldıkça benzerlik artacak, mesafe arttıkça benzerlik azalacaktır.

Bu yaklaşım ideal bir doküman uzayı yaratmada ve doküman sınıflaması yapmada kullanılabilir.

İki doküman arasındaki mesafe, dizin vektörleri arasındaki benzerlik ile ters orantılıdır. En iyi durum, her bir doküman arasında olası en iyi ayırımın olduđu durumdur. Özellikle bir doküman dermesinde iki doküman arasındaki mesafenin belli bir oranda olması arzulanır. Böylelikle bir doküman bir kullanıcı sorgusuna yeterince yakın yerleşimde olduđu zaman erişilmesi garantilenir.

Bu durum, yüksek precision / duyarlılık üretir. Çünkü ilgili dokümanlara erişimi sağlarken ilgisizleri uzak tutabilir. Bir sorguya karşılık birçok ilgili dokümanın uzayın aynı yerel alanında bulunması durumunda ilgili dokümanların çođuna erişilebilir ve ilgisiz olanlar reddedilir. Bu durum hem yüksek anma, hem de yüksek duyarlılık üretir.

Vektör: Vektörler, bilgisayar grafik programları, haritalar, uzaktan algılama, koordinat saptama gibi mühendislik alanlarında kullanılır. Harita mühendisliğinin temel konusudur; gps:global position system, gıs:geographic information system, cođrafik bilgi sistemleri, bilgisayar destekli tasarım ve bilgisayar destekli imalatta kullanılır.

***Hem sayısal hemde yön özelliklerine sahip olan fiziksel nicelikler,
vektörlerle
tems il edilir.***

$$Ben(D_i, D_j) = \frac{a_i \times b_j}{\sqrt{a_1^2 + a_2^2} \times \sqrt{b_1^2 + b_2^2}} = Cos \theta$$