
Yapay zekâ verileri gruplara nasıl ayırır?
www.yzokul.com.tr

Yapay zekâ verileri gruplara nasıl ayırır?



Yapay zekâ verileri gruplara nasıl ayırır? Kısa yanıt, benzer örnekleri bir araya getiren kümeleme adlı bir yöntem kullandığıdır. Kümeleme, bir bilgisayarın grupların ne olduğu söylenmeden veride doğal gruplar bulmasının bir yoludur. Örneğin, bir şarkı listesi verildiğinde kümeleme, kulağa benzer gelen şarkıları gruplayabilir. Bu kitap kümelemeyi sıfırdan başlayarak adım adım açıklar.



Bir grup, basitçe bir araya ait olan şeylerin topluluğudur. Herhangi bir bilgisayar işin içine girmeden önce, öğelerin tek bir yığında toplanmasının ne anlama geldiğini zaten anlıyorsun. Çamaşırları benzer renk yığınlarına ayırmayı ya da kitapları rafta konularına göre gruplamayı düşün. Fikir, bir şeyi paylaşan şeyleri aynı topluluğa koymaktır.



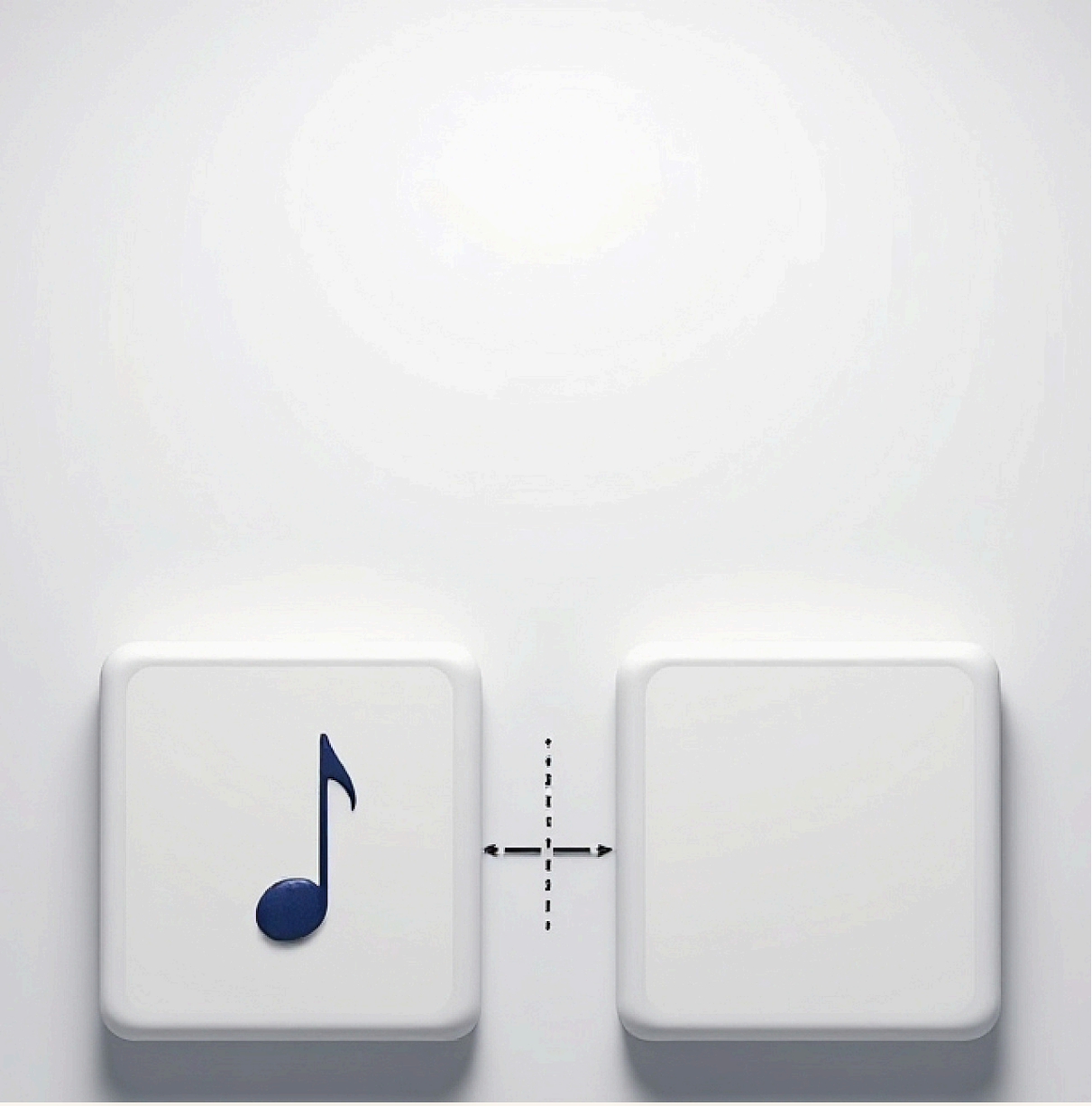
Benzer şeyler bir araya aittir çünkü şeyleri ne kadar benzer olduklarına göre gruplarız. Birçok özelliği paylaşan iki öğe aynı grupta yer alır, farklı olanlar ayrı kalır. Örneğin, elma ve portakal ikisi de meyvedir, bu yüzden bir grupta yer alabilirler; havuç ise sebze grubuna gider. İki öğe ne kadar çok özellik paylaşırsa, birlikte gruplanma olasılıkları o kadar yüksektir.



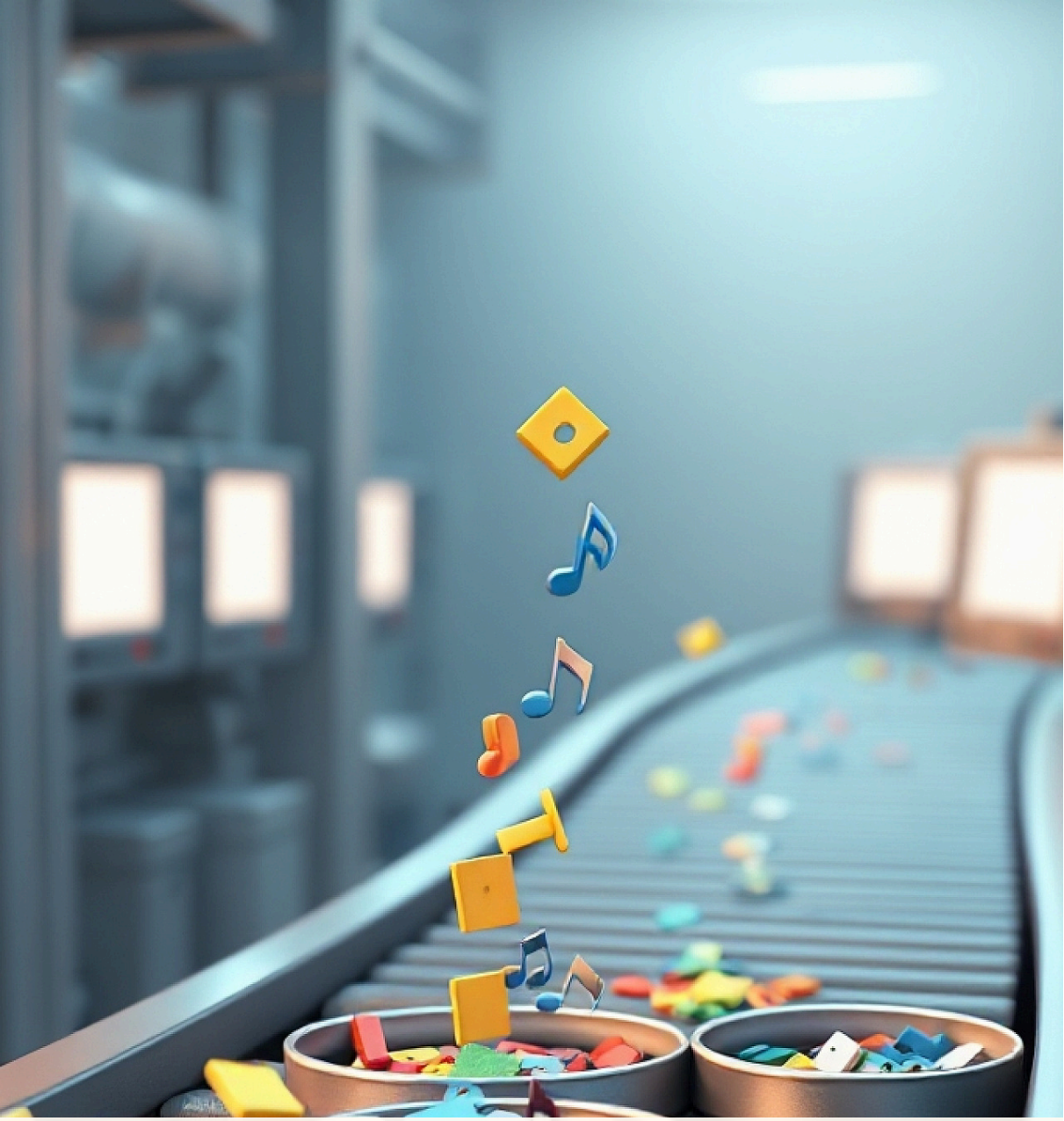
Şarkı örneğiyle tanış: sıralanacak veri olarak büyük bir şarkı yığını hayal et. Her şarkı bir örnektir ve amaç benzer şarkıları aynı gruba yerleştirmektir. Şarkılar tür veya ruh hâli etiketleriyle işaretlenmemiştir; yalnızca bir ses dosyaları topluluğudur. Kümeleme, benzer özellikleri paylaşan şarkı gruplarını bulmaya yardımcı olur.



Özellikler her şarkıyı tanımlar. Bir özellik, bir örnek hakkında ölçülebilir bir ayrıntıdır; örneğin ritim, süre ve ses nitelikleri. Her şarkı için dakikadaki vuruş cinsinden temposunu, saniye cinsinden uzunluğunu ve ortalama ses yüksekliğini ölçebilirsin. Bu sayılar birlikte, şarkıyı tanımlayan bir değerler listesi olan bir özellik vektörü oluşturur. Sistem şarkıları karşılaştırmak için bu özellikleri kullanır.



Benzerliği ölçmek, iki şarkının özelliklerinin ne kadar yakın olduğunu kontrol etmek demektir. Bir benzerlik ölçüsü, iki örneğin ne kadar benzer olduğuna karar vermek için kullanılan bir kuraldır. Örneğin, iki şarkı arasındaki tempo ve süre farkını karşılaştırabilirsin; daha küçük farklar daha yüksek benzerlik anlamına gelir. Yaygın bir ölçü, özellikleri koordinat olarak ele alıp aralarındaki doğrusal uzaklığı hesaplayan Öklid uzaklığıdır. Özellikler ne kadar yakınsa şarkılar o kadar benzerdir.



Kümeleme, benzerlik ölçüsüne göre birbirine yakın olan örnekleri gruplayarak onları bir araya getirir. Benzer ritim, süre veya ses özelliklerine sahip şarkılar aynı kümede yer alır. Kümeleme algoritması örnekleri tekrar tekrar karşılaştırır ve bir grup içindeki öğelerin birbirine, diğer gruplardaki öğelerden daha benzer olması için onları gruplara atar. Örneğin, hızlı ve yüksek sesli şarkılar bir küme oluştururken yavaş ve sessiz şarkılar başka bir küme oluşturabilir.



Önceden hiçbir ad verilmez. Sistemin grup adlarını önceden bilmesi söylenmez; yalnızca özellikleri ve benzerlik kuralını görür. Grupları, özellik değerlerindeki desenleri bularak kendi kendine keşfeder. Örneğin, insanların o kümeye rock müzik dediğini bilmeden benzer tempo ve enerjiye sahip şarkılardan oluşan bir küme bulabilir. Adlar, eğer verilecekse, sonradan gelir.



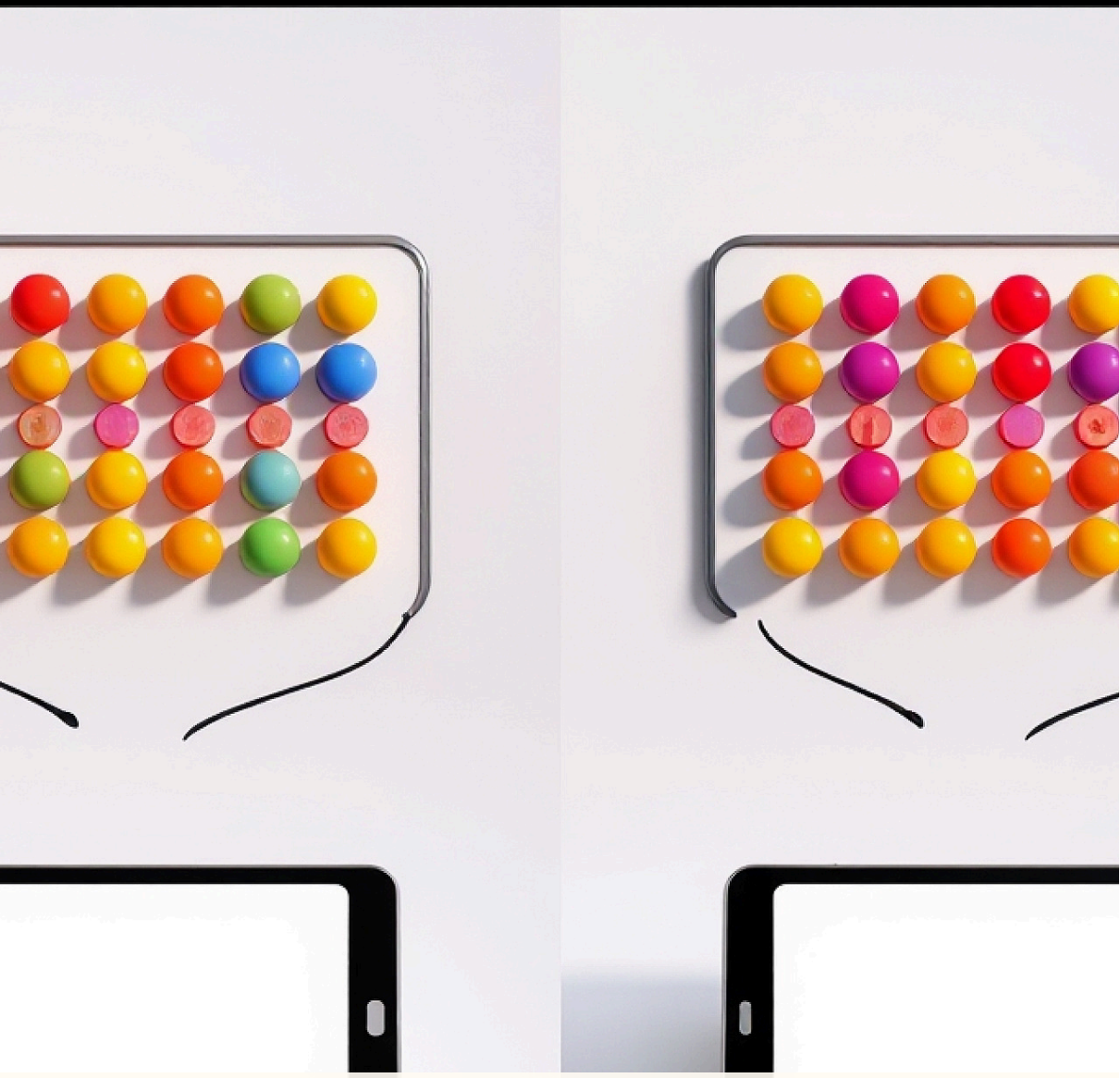
Kümelerin bir resmi yardımcı olur: şarkıları, yakın şarkıların birbirine yakın durduğu bir haritada yerleştirilmiş gibi hayal et. Her şarkı bir noktadır ve noktalar arasındaki uzaklık şarkıların ne kadar benzer olduğunu yansıtır. Kümeler, birbirine yakın yoğun nokta yığınları olarak, daha seyrek bölgelerle ayrılmış biçimde görünür. Bu harita gerçek bir yer değil, özellik uzayını görselleştirmenin bir yoludur.



İnsanlar kümelemeden sonra grupları kontrol eder. Oluşan gruplara bakar ve anlamlı olup olmadıklarını sorar. Benzer şarkılardan oluşan bir grup, ancak insanlar için bir anlam taşıyorsa, örneğin caz gibi kulağa gelen bir grup ya da dans parçalarından oluşan bir grup ise yararlıdır. İnsanlar her kümeden örnekler dinleyip gruplamanın amaçları için anlamlı olup olmadığına karar verebilir.



Özellikleri değiştirmek grupları değiştirir. Şarkıları farklı özelliklerle tanımlarsan gruplar farklı çıkabilir. Örneğin, yalnızca tempoyu kullanmak şarkıları hıza göre gruplayabilirken yalnızca ses yüksekliğini kullanmak onları enerjiye göre gruplayabilir. Seçtiğin özellikler neyin benzer sayılacağını şekillendirir, bu yüzden ilgili özellikleri seçmek yararlı kümeler elde etmek için önemlidir.



Benzerliği değiştirmek grupları değiştirir. Benzerlik ölçüsünü değiştirmek de grupları yeniden düzenleyebilir. Aynı şarkılar bir kural altında bir şekilde, farklı bir kural altında başka bir şekilde kümelenebilir. Örneğin, Öklid uzaklığı ile kosinüs benzerliğini kullanmak, özellik değerlerinin farklı yönlerini vurguladıkları için farklı gruplamalara yol açabilir. Ölçü seçimi sonucu etkiler.



Yaygın hata: önce adlar. Sistemin grup adlarını önceden bilmesi gerektiği doğru değildir. Kümeleme grupları etiketsiz bulur ve adlar yalnızca sonradan insanlar tarafından eklenir. Algoritma yalnızca özelliklerden ve benzerlik ölçüsünden yola çıkarak yapıyı kendi kendine keşfeder. Kümelemenin çalışması için etiketler gerekli değildir.



Yaygın hata: tek doğru yanıt. Tek bir doğru gruplama yoktur. Özellikler ve benzerlik ölçüleri değişebildiği için grupların yapısı da değişebilir. Farklı seçimler, neyi başarmak istediğine bağlı olarak farklı ama yine de geçerli kümeler üretebilir. Kümeleme keşif amaçlıdır ve birden çok gruplama yararlı olabilir.



Yaygın hata: gruplar anlamlıdır. Her küme otomatik olarak yararlı değildir. İnsanlar yine de oluşan grupların kendileri için anlamlı olup olmadığını inceler. Bir küme, özelliklerin veya ölçünün bir yan ürünü olabilir ve gerçek dünyadaki herhangi bir kategoriye karşılık gelmeyebilir. Kümeleri yorumlamak ve doğrulamak için insan yargısı gerekir.



Hatırlanması gerekenler: Yapay zekâ, verileri seçilen özellikler ve bir benzerlik ölçüsü altında benzer olan örnekleri kümeleyerek, grup adları verilmeden gruplara ayırır. İnsanlar grupları kontrol eder ve özellikleri ya da ölçüyü değiştirmek grupları değiştirebilir. Kümeleme, verideki yapıyı keşfetmek için güçlü bir araçtır, ancak düşünceli seçimler ve insan yorumu gerektirir.