

Makine Öğrenmesi algoritmalarının türleri , yaklaşımlarına , veri türüne göre ve problemin türlerine göre farklılık gösterir. Bu farklı yaklaşımlara girmeden önce bazı kavramları irdeleyelim.

1.Öznitelik Vektörleri:

Öznitelik, makine öğrenmesinde ve örüntü tanımada, analiz edilen veri yığınının veya setinin anlaşılır , anlamlı , ölçülebilir bir özelliğidir.

Örnek olarak bir banka müşterisine ait bir veri seti olsaydı, üyelik senesi , hesap türü, yaş, cinsiyeti, ödeme alışkanlıkları bunlar birer öznitelik olabilirdi. Veya bir e-ticaret sitesi için geliştirilecek bir öneri motoru düşünelim. Bu motoru eğitmek için veri setimizdeki kullanıcılara ait gezilen kategoriler , alınan ürünler , satın alma sıklığı, cinsiyeti gibi veriler öznitelik olabilirdi. Başka bir örnekle bir çiçek türüne ait bir veri setinde o çiçeğe ait yetiştiği bölge ve rengi , yaprak genişliği ve uzunluğu öznitelikleridir.

Öznitelik vektörleri ise bir nesnenin özniteliklerini matematiksel olarak , kolayca analiz edilebilir olması için vektörler halinde temsil edilmesidir.

Bilgisayarlar 0,1'lerden matrislerden ve vektörlerden anlar. Verilerimizi bilgisayarın anlayacağı hale getirmemiz gerekir. Öznitelik vektörü nesne hakkında birden çok özellik içerir. Çok boyutlu veya tek satır , tek sütun şeklinde de olabilir.

Örnek bir renk içinde ne kadar kırmızı, mavi ve yeşil olduğu ile tanımlanabilir. Bunu bir öznitelik vektörü haline getirirsek renk=[Red,Green,Blue] olur. Öznitelik vektörleri özellikle görüntü işleme verileri için oldukça kullanışlıdır.

2.Eğitim Modelleri

Bir makine öğrenimi modeli kendisine verilmiş veri setinden örüntü çıkarımı yapabilen ve karar verebilen modellerdir. Bu modellerin eğitiminde çok fazla veri gerekir. Örnek olarak denetimli öğrenmede, algoritma birçok örneği inceler ve doğruluğu en yüksek olan modeli oluşturur. Bu iyi eğitilmiş bir modeldir. Eğitim modellerinde amaç düşük kayıplı bir sonuca ulaşabilmektir. Bu eşik değeri bulmada ortalama kareler hatası(MSE) kullanılır.

Ortalama Kareler Hatası:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Burada N, veri kümesindeki örnek sayısı.

y_i = gerçek değerler

\hat{y}_i = modelin tahmin ettiği değer

Örnek $y_i=1,2,3,4,5$, $\hat{y}_i=1,2,4,3,5$ (modelin tahminleri) $N=5$

$$MSE=((1-1)^2 + (2-2)^2 + (3-4)^2 + (4-3)^2 + (5-5)^2) / 5$$

$$=(0+0+1+1+0)/5$$

$$=2/5=0.4 \quad MSE=0.4 \text{ düşük kayıplı bir tahmin işlemi}$$

3.Entropi(Bilgi Kazancı):

Teorik olarak rassal bir değişkenin belirsizlik ölçütüdür. Beklenen değeridir.Eşit olasılıklı durumlar yüksek belirsizliği temsil eder.

Makine öğrenmesinde entropi, belirli bir veri kümesi veya sistemdeki düzensizlik veya belirsizlik düzeyini ölçer. Bir veri kümesindeki bilgi miktarını ölçen bir ölçümdür ve genellikle bir modelin kalitesini ve doğru tahminler yapma yeteneğini değerlendirmek için kullanılır.

ML 'de entropi ne kadar düşükse modelin karar verme/tahmin etmesi o kadar isabetli olur

Formül:

$$Entropy(p) = - \sum_{i=1}^N p_i \log_2 p_i$$

P_i olasılık değeri , N ise tüm olasılık miktarıdır.

Play Golf	
Yes	No
9	5



$$\begin{aligned} Entropy(PlayGolf) &= Entropy(5,9) \\ &= Entropy(0.36, 0.64) \\ &= -(0.36 \log_2 0.36) - (0.64 \log_2 0.64) \\ &= 0.94 \end{aligned}$$

$N=2$ iki olası durum var
Yes ya da No

$$P_i \text{ No için} = 5/14 = 0.3571$$

$$P_i \text{ Yes için} = 9/14 = 0.6428$$

4.Eksik Veri:

Veri setinde bazı veriler eksik/kayıp olabilir bu durumda iki yol izlenir;

- 1.Eksik özelliklere sahip veriler tamamen veri setinden çıkarılır
- 2.Eksik veriler doldurularak model eğitilir

Ya da eksik verinin olduğu sütunun ortalama değerinde eksik veriler yerine yazılabilir.

Eksik veriler ile mücadele edilmezse modelimizin doğruluk değerini düşürebilir.