

Istatistik - Ders 1

Bu notlar makine öğrenimi, veri madenciliği gibi konularda gerekli olasılık ve istatistik bilgisini paylaşmak için hazırlanıyor. Notlarda olasılık ve istatistik aynı anda anlatılacak, ve uygulamalara ağırlık verilecek.

Orneklem Uzayı (Sample Space)

Orneklem uzayı Ω bir deneyin mümkün tüm olasılıksal sonuçların (outcome) kümesidir. Eğer deneyimiz ardi ardına iki kere yazi (T) tura (H) atıp sonucu kaydetmek ise, bu deneyin mümkün tüm sonuçları şöyledir

$$\Omega = \{HH, HT, TH, TT\}$$

Sonuçlar ve Olaylar (Outcomes and Events)

Ω içindeki her nokta bir sonuctur (outcome). Olaylar Ω 'nin herhangi bir alt kümesidir ve sonuçlardan oluşurlar. Mesela üstteki yazi-tura deneyinde “iki atisin icinden ilk atisin her zaman H gelmesi olayi” böyle bir alt kümedir, bu olaya A diyelim, $A = \{HH, HT\}$.

Ya da bir deneyin sonucu ω fiziksel bir ölçüm , diyelin ki sıcaklık ölçümü. Sıcaklık \pm , reel bir sayı olduğuna göre, $\Omega = (-\infty, +\infty)$, ve sıcaklık ölçümünün 10'dan büyük ama 23'ten küçük ya da esit olma “olayı” $A = (10, 23]$. Koseli parantez kullanildi cunku sinir degerini dahil ediyoruz.

Ornek

10 kere yazi-tura at. A = “en az bir tura gelme” olayi olsun. T_j ise j 'inci yazi-tura atisinda yazi gelme olayi olsun. $P(A)$ nedir?

Bunun hesabi için en kolayi, hic tura gelmeme, yani tamamen yazi gelme olasiligini, A^c 'yi hesaplamak, ve onu 1'den cikartmaktır. c sembolu “tamamlayici (complement)” kelimesinden geliyor.

$$\begin{aligned} P(A) &= 1 - P(A^c) \\ &= 1 - P(\text{hepsi yazi}) \\ &= 1 - P(T_1)P(T_2)...P(T_{10}) \\ &= 1 - \left(\frac{1}{2}\right)^{10} \approx .999 \end{aligned}$$

Rasgele Degiskenler (Random Variables)

Bir rasgele degisken X bir eslemedir, ki bu esleme $X : \Omega \rightarrow \Re$ her sonuc ile bir reel sayi arasindaki eslemedir.

Olasilik derslerinde bir noktadan sonra artik ornekleme uzayindan bahsedilmez, ama bu kavramin arkalarda bir yerde her zaman devrede oldugunu hic aklimizdan cikartmayalim.

Ornek

10 kere yazi-tura attik diyelim. VE yine diyelim ki $X(\omega)$ rasgele degiskeni her ω siralamasinda (sequence) olan tura sayisi. Iste bir esleme. Mesela eger $\omega = HHTHHTHHTT$ ise $X(\omega) = 6$. Tura sayisi eslemesi ω sonucunu 6 sayisina esledi.

Ornek

$\Omega = \{(x, y); x^2 + y^2 \leq 1\}$, yani kume birim cember ve icindeki reel sayilar (unit disc). Diyelim ki bu kumeden rasgele secim yapıyoruz. Tipik bir sonuc $\omega = (x, y)$ 'dir. Tipik rasgele degiskenler ise $X(\omega) = x$, $Y(\omega) = y$, $Z(\omega) = x + y$ olabilir. Goruldugu gibi bir sonuc ile reel sayi arasinda esleme var. X rasgele degiskeni bir sonucu x 'e eslemis, yani (x, y) icinden sadece x 'i cekip cikartmis. Benzer sekilde Y, Z degiskenleri var.

Toplamsal Dagilim Fonksiyonu (Cumulative Distribution Function -CDF-)

Tanim

X rasgele degiskeninin CDF'i $F_X : \Re \rightarrow [0, 1]$ tanimi

$$F_X(x) = P(X \geq x)$$

Eger X ayriskal ise, yani sayilabilir bir kume $\{x_1, x_2, \dots\}$ icinden degerler aliyorsa olasilik fonksiyonu (probability function), ya da olasilik kutle fonksiyonu (probability mass function -PMF-)

$$f_X(x) = P(X = x)$$

Bazen f_X , ve F_X yerine sadece f ve F yazariz.

Tanim

Eger X surekli (continuous) ise, yani tum x 'ler icin $f_X(x) > 0$, $\int_{-\infty}^{+\infty} f(x)dx =$

1 olacak sekilde bir f_X mevcut ise, o zaman her $a \leq b$ icin

$$P(a < X < b) = \int_a^b f_X(x)dx$$

Bu durumda f_X olasilik yogunluk fonksiyonudur (probability density function -PDF-).

$$F_X = \int_{-\infty}^x f_X(t)dt$$

Ayrica $F_X(x)$ 'in turevi alinabildigi her x noktasinda $f_X(x) = F'_X(x)$ demektir.

Dikkat! Eger X surekli ise o zaman $P(X = x) = 0$ degerindedir. $f(x)$ fonksiyonunu $P(X = x)$ olarak gormek hatalidir. Bu sadece ayriksal rasgele degiskeninler icin isler. Surekli durumda olasilik hesabi icin belli iki nokta arasinda integral hesabi yapmamiz gereklidir. Ek olarak PDF 1'den buyuk olabilir, ama PMF olamaz. PDF'in 1'den buyuk olabilmesi integrali bozmaz mi? Unutmayalim, integral hesabi yapiyoruz, noktasal degerlerin 1 olmasi tum 1'lerin toplandigi anlamina gelmez. Bakiniz Entegralleri Nasil Dusunelim yazimiz.