

Ata Atas
040230771

e) Hesapladığınız ortalama ve standart sapma değerleri a) da verilerle aynı mıdır? Değil ise nedenini yorumlayınız.

a) sıklıkta verilen ortalama ve standart sapma değerleri:

sınıf 1 \rightarrow ort: 2 & std: 1

sınıf 0 \rightarrow ort: 0 & std: 1

Berlin kodum (Standart mat hesaplaması ile bulduklarım "custom"; numpy kütüphanesinin kendi metodlarıyla bulduklarım "np" şeklinde adlandırılmıştır.)

sınıf 1 \rightarrow custom ort: 2,01889 & custom std: 1,1585999

np ort: 2,01889 & np std: 1,1556

sınıf 0 \rightarrow custom ort: -0,0013443 & custom std: 1,1536

np ort: -0,0013443 & np std: 1,1507

Görüldüğü üzere 2 farklı method ile bulunan değerler de soruda verilen popülasyon değerlerine oldukça yakındır. Aradaki küçük farklılaşmanın sebebi, doğrudan gauss dağılımı üretmek yerine, sabit aralıklarla ürettiğim rastgele evliği, gauss dağılımına uyduracak exp fonksiyonuna manuel girmiş olmam ve popülasyonun tamamı için verilen değerleri 200 elementli bir örnekleme grubundan elde etmeye çalışıyor olmamdır.

f) Gauss dağılımına uygun sayı üretmenin algoritmasını matematiksel olarak açıklayınız.
> Dijital 0 ve 1 sinyallerini veren gerilim değerlerinde eşit uzaklıkları, sınıf seçerek python random kütüphanesini kullanarak random.uniform() komutuyla verdiğim değerler arasından float sayılar oluşturdum (0V ve 2V için 2'er birim uzaklığa sınıf seçtim \rightarrow (2,2)/(0,4) gibi). oluşturduğum float sayıların pdf (probability distribution function)'ini bulabilmek için gauss dağılımı olasılığını veren formülü kendim yazarak dağılımın kendi bulduğum ortalama ve standart sapma değerlerini de fonksiyona girip dağılım değerlerini liste şeklinde verdim ve gauss dağılımı elde ettim ($\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ kullandım)

1)

1) Tasarladığınız sınıflandırıcının performansı yeterli midir? Yorumlayınız.

Accuracy: 0,73 → Sınıflandırıcı %73 oranında doğru tahmin yapar.

Precision: 0,725 → Pozitif yanı sınıfl tahmininde %72,5 oranında doğrudur.

Recall: 0,732 → Gerçekten pozitif olan değerlerin %73'ünü doğru bilir.

F1 score: 0,729 → precision & recall arasında %73 oranında değerli.

%73 yeterli midir?

Basit günlük sensör sınıflandırmasında → kabul edilebilir.

Endüstriyel kontrol ve güvenlik sistemlerinde → genellikle %95 doğruluk hedefleri ve yetersizdir.

Akademik demo/konsept geçerliliği kanıtı → konseptin çalıştığına işaret olabilir.

Medikal sınıflandırma → yetersiz, ciddi sonuçlara sebep olabilir.

Kısaca modelin yeterlilik seviyesi yapılmak istenen işe göre nispetlidir. Ama olasılıksal ve matematiksel olarak kıyası nedir?

Sınıf 0: $X_0 \sim N(\mu_0=0, \sigma_0=1)$ sınıf 1: $X_1 \sim N(\mu_1=2, \sigma=1)$

Önsel olasılıklar eşittir yani → $\pi_0 = \pi_1 = \frac{1}{2}$ ve varyans da eşittir yani optimal karar noktası orta noktadır. → $\frac{\mu_0 + \mu_1}{2} = 1 = x$

Bayes Hatası (minimum): $P_e = \pi_0 P(X_0 > x) + \pi_1 P(X_1 < x) = \frac{1}{2} P(X_0 > x) + \frac{1}{2} P(X_1 < x)$
Simetriden ve eş önsel olasılıklardan bu ikisi eşittir. $\equiv P(X_0 > x)$

Tablo kullanabilmek için Z tablosuna uygun dönüşüm yaparın ve standardize ettim:

$X_0 \sim N(\mu_0=0, \sigma_0=1) \rightarrow X_0 = \mu_0 + \sigma_0 Z \rightarrow Z = \frac{X_0 - \mu_0}{\sigma_0} \rightarrow P_e = P(X_0 > 1)$

$P_e = P(\mu_0 + \sigma_0 Z > 1) \rightarrow P_e = P(Z > \frac{1 - \mu_0}{\sigma_0}) \rightarrow P_e = P(Z > 1)$
CDF (kümülatif dağılım fonk.) cinsinde yazarsak $\Phi(z) = P(Z \leq z)$

$P(Z > 1) + P(Z \leq 1) = 1$ ise $P(Z > 1) = 1 - \Phi(1)$ standart normal dağılım tablosundan $\Phi(1) = 0,8413$ gelir ve sonuç olarak $P_e = P(Z > 1) = 1 - 0,8413 = 0,1587 = \boxed{\%15,87}$

yani beirim 0,73'lük accuracy değerimden gelen %27 hata ideal sınıflandırıcının hatasına göre 11-12 puan daha fazla. Sisteme göre nispetli bakarsak $(\frac{27-15,87}{27}) \rightarrow \%41,1$ 'lik bir hata iyileştirmesi yapılabilir ve daha iyisi dağılımların kesişmesinde istene mümkün değildir. İdeal hatanın neredeyse 2 katı bir hata mevcuttur ama sisteme göre %27 hata yeterliliği görülebilir.