

Açıklama ve uyarılar: Ödev *el yazısı ile hazırlanacak* ve yanıtlar pdf dosyasına çevrildikten sonra en geç 27 Kasım tarihinde uzak.etu.edu.tr sistemi üzerinden teslim edilecektir. E-posta vb farklı yollarla gönderilen ya da geç teslim edilen ödevler dikkate alınmayacaktır.

Sorular

1. Türkiye'ye ait il bazlı bir sınıflandırma çözümlemesinde X = ildeki firma sayısı (bin adet) verilerini kullanarak Y = ilde sanayi odası olup olmayacağını (yok ise 0, var ise 1) kestirmek istiyoruz. Buna ilişkin lojistik regresyon tahminleri aşağıda verilmiştir:

Call:

```
glm(formula = oda ~ firma, family = binomial, data = veriseti)
```

Deviance Residuals:

```
      Min       1Q   Median       3Q      Max
-3.2235  -0.3602  -0.3197  -0.2794   2.0263
```

Coefficients:

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.3083      0.6350  -5.210 1.89e-07 ***
firma         1.8571      0.6218   2.987 0.00282 **
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 63.766  on 78  degrees of freedom
Residual deviance: 41.518  on 77  degrees of freedom
AIC: 45.518
```

- (a) Tahmin çıktılarını kullanarak 1737 adet firmaya sahip bir ilde sanayi odası bulunma olasılığını hesaplayınız. (*Dikkat:* Katsayı tahminlerini ondalık virgülden sonra 1. basamağa yuvarlayınız ve $e = 3$ kabul ediniz.)
- (b) Model tahminlerine göre 1500 adet firmaya sahip bir ilin sınıfı ne olur?
2. Birinci soruda tahmin edilen sınıflandırma modeline ait karışıklık matrisi aşağıdaki gibidir:

		Kestirilen Durum		
		Var	Yok	Toplam
Gerçek Durum	Var	5	8	13
	Yok	1	67	68
Toplam		6	75	81

- (a) Sınıflandırmaya ait *doğru pozitif* oranı nedir?
- (b) Sınıflandırmaya ait *yanlış pozitif* oranı nedir?
- (c) Sınıflandırmaya ait *doğru negatif* oranı nedir?

TOBB - Ekonomi ve Teknoloji Üniversitesi
İKT457 – Ekonomi ve Finans için Yapay Zeka I, Üçüncü Ödevi

3. Birinci soruda kullanılan “veriseti” adlı R veri çerçevesi objesine ait değişken adları aşağıda verilmiştir:

```
> names(veriseti)
[1] "oda"    "firma"
>
```

- (a) “oda” adlı nitel değişkeni “firma” değişkenini kullanarak *lojistik* model ile sınıflandıracak R komut(lar)ını yazınız.
- (b) “oda” adlı nitel değişkeni “firma” değişkenini kullanarak *doğrusal DÇ* modeli ile sınıflandıracak R komut(lar)ını yazınız.
- (c) “oda” adlı nitel değişkeni “firma” değişkenini kullanarak *K-EK* ($k = 3$) modeli ile sınıflandıracak R komut(lar)ını yazınız.
4. Doğrusal ve ikinci derece diskriminant çözümlemesine yönelik olarak aşağıdaki soruları yanıtlayınız:
- (a) Bayes karar sınırı eğer doğrusal ise *eğitim veri setinde* doğrusal DÇ’nin mi yoksa ikinci derece DÇ’nin mi daha iyi sonuç vermesini beklersiniz? Kısaca açıklayınız.
- (b) Bayes karar sınırı eğer doğrusal-dışı ise *test veri setinde* doğrusal DÇ’nin mi yoksa ikinci derece DÇ’nin mi daha iyi sonuç vermesini beklersiniz? Kısaca açıklayınız.
- (c) X değişkenleri sayısı arttıkça ikinci derece DÇ çözümlemesinin doğrusal DÇ’ye göre giderek daha iyi sonuç vermesini mi yoksa sonuçların kötüleşmesini mi beklersiniz? Kısaca açıklayınız.
- (d) Gözlem sayısı arttıkça ikinci derece DÇ çözümlemesinin sade Bayes’e göre giderek daha iyi sonuç vermesini mi yoksa sonuçların kötüleşmesini mi beklersiniz? Kısaca açıklayınız.
5. Bahis oranı kavramına ilişkin olarak aşağıdaki soruları yanıtlayınız:
- (a) Bir grup firmaya yönelik iflas bahis oranı eğer 1’e 7 ise bu firmaların ortalamada yüzde kaçının iflas etmesini beklersiniz?
- (b) Bir grup firmaya yönelik iflas bahis oranı eğer 7’ye 1 ise bu firmaların ortalamada yüzde kaçının iflas etmesini beklersiniz?