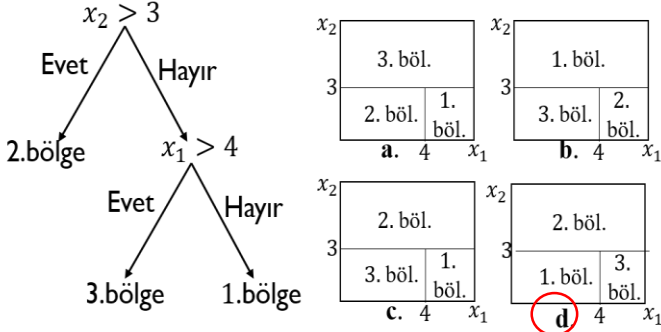


**A -Sivas Cumhuriyet Üniversitesi Müh. Fakültesi
Bilgisayar Müh. Böl. 2022-2023 Güz Dönemi Bil4211
Veri Madenciliği Final Soruları**

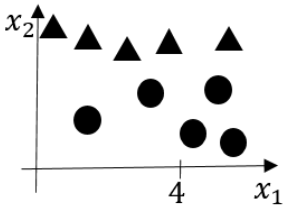
1. Gruplama (clustering) ile aşağıdaki ifadelerden hangisi yada hangileri doğrudur?

- I. Grup sayısı sabittir.
II. Her grupta aynı sayıda örnek olmak zorundadır.
III. Gruplandırma, sınıflandırmanın ön işlemidir; yani gruplamadan sonra sınıflandırmaya geçilir.
a) Hepsisi b) Hiçbiri c) Yalnız III d) I ve III e) II ve III

2. Aşağıda verilen karar ağacının öklid uzayında ayırdığı bölgeler hangi şıkta doğru olarak verilmiştir?



e) Hiçbiri



3. Yanda x_1 ve x_2 özellikleriyle ifade edilen üçgen ve yuvarlak kullanılarak sembolize eden iki farklı sınıfa ait eğitim örnekleri görülmektedir. Buna göre

aşağıdakilerden hangisi yada hangileri doğrudur?

- I. Tek başına x_1 özelliği tek başına x_2 özelliğine göre sınıfları belirlemede daha ayırt edicidir.
II. x_1 değeri 4'ten büyük olan bir test örneği k-en yakın komşu ile sınıflandırılmak istenildiğinde k, 3 alınırsa yuvarlak sınıf olarak sınıfı tahmin edilir.
III. Sınıflandırmak istediğimiz bir test örneğinin x_1 değeri x_2 değerinden daha fazla ise yuvarlak sınıfa ait olması daha olasıdır.

a) Hepsisi b) Hiçbiri c) Yalnız II d) I ve II e) II ve III

4. Aşağıdakilerden hangisi en yakın komşu algoritmasının en önemli dezavantajıdır?

- a) Düşük sınıflandırma performansı
b) Uç örnekler (outlier'lar) a karşı hassasiyet
c) Az hafızaya gerek duyması
d) Fazla hesaplama gerektirmesi
e) Hiçbiri

5.

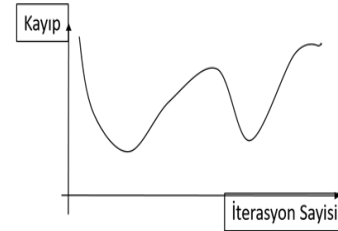
Gerçek Sınıflar	Tahmin Edilen Sınıflar	
	Pozitif	Negatif
Pozitif	4	0
Negatif	3	5

Yukarıda gösterilen karışıklık matrisine göre aşağıdakilerin kaç doğrudur?

- * Test setinde 12 örnek vardır.
* Test setinde 7 pozitif, 5 negatif örnek vardır.
* Pozitif örneklerin tamamı doğru sınıflandırılmıştır.
* Kesinlik %66'dır.
* Pozitif örnekleri sınıflandırmak daha zordur.
a) 5 b) 4 c) 3 d) 2 e) 1

6. Lojistik regresyonla ilgili ifadelerden aşağıdakilerden kaç tanesi her zaman doğrudur?

- I. Yalnızca iki sınıfın olduğu sınıflandırma için kullanılır
II. Özelliklere karşılık gelen w_0, w_1, \dots, w_d katsayılarına sahip olduğumuzda, bir test örneğinin pozitif sınıfa ait olma olasılığını direkt hesaplayabiliriz.
III. Naive Bayes ve en yakın komşu algoritmasına göre daha iyi tahmin yapar.
a) Yalnız II b) I ve II c) II ve III d) Yalnız e) Hepsisi



7. Yandaki bir gradient descent (dereceli azalma) uygulamasına ait iterasyon sayısının artmasıyla oluşan kayıp (loss) miktarını gösteren bir grafik verilmiştir. Bu

grafığın bu şekilde çıkmasına aşağıdaki nedenlerden hangisi neden olabilir?

- a) Kötü bir başlangıç değeri alınmıştır.
b) Öğrenme oranı çok küçük seçilmiştir.
c) Öğrenme oranı çok büyük seçilmiştir
d) Kayıp fonksiyonu türevlenebilir değildir
e) Kısmi türev yanlış alınmıştır.

8. ve 9. soruyu aşağıdaki eğitim setine göre yapınız.

Araba Sahibi	Ev Sahibi	Evli	Krediyi Ödenmiş mi?
Evet	Evet	Hayır	+
Hayır	Hayır	Evet	+
Hayır	Hayır	Hayır	-
Evet	Hayır	Evet	-

8. Aşağıdaki ifadelerden hangisi yada hangileri doğrudur?

I. Arabasi ve evi olmayip evli olan birinin alacağı krediyi ödeyip ödeyemeyeceği kararı en yakın komşu ($k=1$) algoritmasıyla yada Naive Bayes algoritmasıyla kullanılarak karar verilirse, iki durumda da aynı sonuca ulaşılır.

II. Naive Bayes kullanılarak bir kişinin alacağı krediyi ödeyip ödeyemeyeceğine karar verilirse, pozitif ve negatif sınıfın kendi başına olabilirliğini, yani bu sınıfların herhangi bir veri olmadan olma olasılıklarını dikkate almak bir fark oluşturmaz.

III. Arabasi ve evi olan evli birinin alacağı krediyi ödeme olasılığı ile ödememe olasılığı aynıdır.

a) Yalnız II b) Yalnız I c) Hiçbiri d) I ve II e) II ve III

9. Arabasi ve evi olmayip evli olan birinin alacağı krediyi ödeme olasılığı Naive Bayes kullanılarak hesaplanırsa bu olasılık ne olur?

a) 1 b) 1/2 c) 1/3 d) 1/4 e) 1/5

10. Aşağıdakilerden hangisi denetimsiz (unsupervised) öğrenmeye bir örnektir?

a) k-en yakın komşu b) Lineer Regresyon c) Karar Ağacı d) Gradient Descent e) Kümeleme

A	B	C
0	1	0
1	0	1
1	1	0
1	0	0
1	1	0
0	0	0

11. Yanda A,B ve C adlı üç kategorik özellekle gösterilen altı adet örnekten oluşan bir veri seti görülmektedir. Buna göre A, B ve C özelliklerinin entropilerinin (H_A, H_B, H_C) büyükten küçüğe sıralanması hangi şıkta doğru olarak verilmiştir?

a) $H_C > H_A > H_B$

b) $H_B > H_A > H_C$ c) $H_A > H_C > H_B$

d) $H_C > H_B > H_A$ e) $H_B > H_C > H_A$

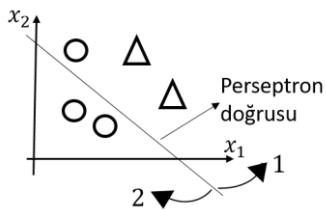
12. Aşağıdakilerden hangisi yada hangileri boyut azaltmanın motivasyonlarından biridir?

I. Sınıflandırıcıların performansını artırmak

II. Örnekleri/objeleri görüntülenmesini sağlamak

III. Örnek/obje sayısını azaltmak

a) Hepsi b) Yalnız I c) I ve II d) I ve III e) II ve III



13. Yanda x_1 ve x_2 özellikleriyle ifade edilen; yuvarlak veya üçgen sınıflarından birine ait olan beş örnek gösterilmiştir. Bu veri

üzerinde çalışan perseptron algoritmasının bu iki sınıfı ayırmak için bulduğu (perseptron) doğrusu şeklindeki gibidir. Buna aşağıdakilerden hangisi doğrudur?

a) Perseptron bir iterasyon daha çalışırsa, perseptron

doğrusu 1 yönünde bir miktar döner ve sonraki tüm iterasyonlar boyunca bu doğru sabit kalır.

b) Perseptron bir iterasyon daha çalışırsa, perseptron doğrusu 2 yönünde bir miktar döner ve sonraki tüm iterasyonlar boyunca bu doğru sabit kalır.

c) Perseptron kaç iterasyon daha çalışırsa çalışsın doğrunun yeri değişmez.

d) Perseptron kaç iterasyon daha çalışırsa çalışsın doğru sürekli yer değiştirir; hiçbir zaman sabit kalmaz.

e) Hiçbiri

P1
P2
P3
P4
P5

14. Diyelimki veri setini yandaki gibi 5 parçaya ayırıp 5-katlı çapraz doğrulama ile sınıflandırma algoritmasının başarısını ölçmek istiyoruz. Buna göre aşağıdakilerden hangisi yada hangileri doğru olur?

I. Sınıflandırıcı 5 defa eğitilir.

II. Her bir parçadan gelen kesinlik (accuracy) birbirinin aynısı olmak zorundadır.

III. Her parçada tüm sınıflara ait örnekler bulunmak zorundadır.

a) Hepsi b) Hiçbiri c) Yalnız III d) Yalnız I e) I ve II

15. Aşağıdaki ifadelerden hangisi yanlıştır?

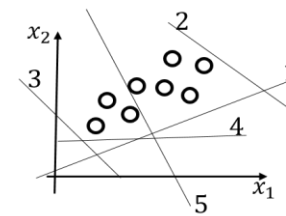
a) Lojistik regresyon reel sayı tahmininde kullanılamaz.

b) Kümelemede ortaya çıkacak küme/grup sayısı sabit değildir, kullanılan algoritmaya göre değişebilir.

c) Aşırı öğrenme (overfitting) eğitim setinin aşırı öğrenilmesiyle oluşur.

d) Veri setindeki her örneğin bir kümeye dahil olma zorunluluğu yoktur.

e) Hiçbiri



16. Yanda x_1 ve x_2 özellikleriyle ifade edilen 8 adet örnek görülmektedir. Bu örneklerin boyutunu PCA kullanarak bir boyuta indirmek istersek PCA bize

kaç nolu doğruyu bu projeksiyon doğrusu olarak verir?

a) 1 b) 2 c) 3 d) 4 e) 5

17. Kaç tanesi yanlıştır?

I. Veri madenciliği çalışmalarının öngörülebilir/tahmin edilebilir sonuçlar vermesi beklenir.

II. Karar ağaçları ile oluşturulan saf bölgelerde entropi yüksektir.

III. İyi bir özellik, tüm örnekler için mümkün olduğunca aynı değeri almalıdır.

IV. Validasyon seti, sınıflandırıcının ne kadar başarılı olduğunu ölçmek için kullanılır.

a) 0 b) 1 c) 2 d) 3 e) 4