

Ad-Soyad:

1.Öğretim 2. Öğretim No:

Sivas Cumhuriyet Üniversitesi Mühendislik Fak. Bilgisayar Müh. Bölümü. Bil4211 2018-Güz Final Sınavı

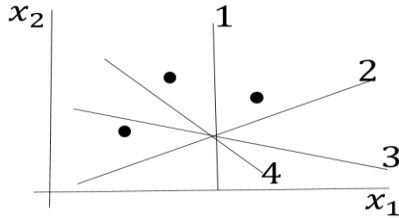
1. Aşağıdakilerden hangisi veri madenciliğinde karşılaşılan güçlüklerden biri değildir?

- a. Sürekli veri akışının olması.
b. Verinin çok olması.
c. Verinin çeşitli kaynaklardan elde edilmesi.
d. Verinin toplandığı alanın bilinmesi gerekliliği.

2. Doğru olan seçeneği işaretleyiniz.

- a.** Kategorik tipteki özelliklerle daha az sayıda işlem yapılabileceğinden daha değerlidir.
b. Tutarsız veri veriye çeşitlilik katması açısından önemlidir.
c. Bir veri setinde boyut arttıkça objeler birbirine yakınlaşır.
d. Özellikleri filtreleme yöntemiyle seçersek özellikler arası korelasyonu dikkate almış oluruz.

3.

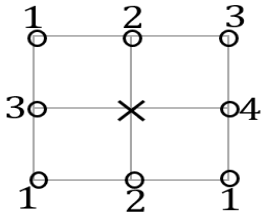


Yukarıda nokta ile gösterilen üç eğitim örneği PCA kullanılarak iki boyuttan tek boyuta indirilmek istenmektedir. Buna göre PCA'nın bulacağı doğru numaralandırılmış doğrulardan hangisi olur?

- a. 1 **b.** 2 c. 3 d. 4

4. Aşağıda iki boyutlu uzayda verilen test örneği x ile; eğitim örnekleri o ile gösterilmiştir. Bu örneklerin ait oldukları sınıflar belirtilmiştir. Buna göre eğer test örneğini $k=3$ olacak şekilde en yakın komşu algoritması ile sınıflandırsak sınıfı ne olur?

- a. 1 **b.** 2 c. 3 d. 4

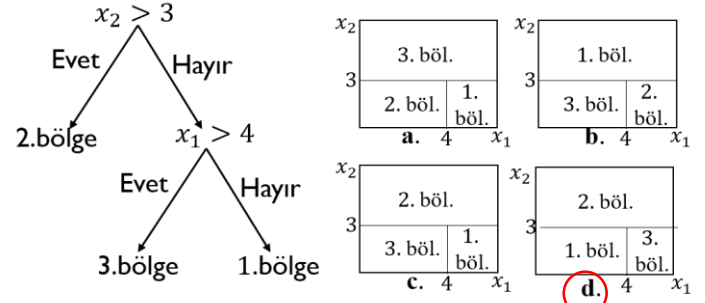


5. Perceptron algoritması ile ilgili aşağıda yazılanlardan hangisi yada hangileri doğrudur?

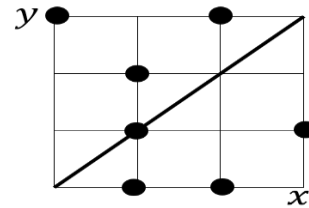
- I. Tembel bir öğrenme algoritmasıdır.
II. Eğitim örneklerinin tamamı doğru sınıflanıncaya kadar algoritma sonlanmaz.
III. Yanlış sınıflandırılmış her örnek için w katsayıları güncellenir.

- a. Yalnız III b. I ve II **c.** II ve III d. I, II, III.

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



7. Aşağıda tek boyutlu (yalnız x) yedi eğitim örneği, bu örneklerin y değerleri ve bu örnekler için oluşturulan regresyon doğrusu gösterilmiştir. Bu regresyon doğrusunun yol açtığı $\frac{1}{2m} \sum_{i=1}^m (f(x^i) - y^i)^2$ formülü ile verilen ortalama maliyet ne olur? (m örnek sayısı)



- a. 2 b. 1.5 c. 0.71 **d.** 1.42

8. Gradient descent ile ilgili aşağıdaki bilgilerin hangisi yanlıştır?

a. Gradient descent bir maliyet fonksiyonunda minimum değeri veren değişkenleri bulmamızı sağlar.

b. Gradient descent bir lokal minimum takılmışsa adım büyüklüğünü değiştirmek fayda sağlamaz.

c. Gradient descent maliyet fonksiyonunu düşürmeyi durdurmuşsa o noktada türev 0 olduğu içindir.

d. Kümeleme analizinde küme merkezlerinin iteratif olarak bulunmasında gradient descent kullanılabilir.

9. Aşağıda tek bir özelliğe sahip dört hayvan ve bu hayvanların etinin güvenli olup olmadığı verilmiştir.

Renk	Sınıf
Sarı	Güvenli
Sarı	Güvensiz
Kahverengi	Güvenli
Sarı	Güvenli

Sarı bir hayvanın güvenli olma olasılığı nedir?

- a. $\frac{3}{4}$ b. $\frac{1}{3}$ c. $\frac{2}{3}$ d. $\frac{1}{2}$

10. ve 11. soruları bir test setinden elde edilen aşağıdaki karışıklık matrisine göre çözünüz.

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

10. Aşağıdakilerin kaç doğrudur?

- * Test setinde 12 örnek vardır. +
- * Test setinde 7 pozitif, 5 negatif örnek vardır.
- * Pozitif örneklerin tamamı doğru sınıflandırılmıştır. +
- * Kesinlik %66'dır.

- a.4 b.3 c.2 d.1

11. Aşağıdakilerin kaç yanlıştır?

- * Gerçek negatif oran $\frac{3}{7}$ 'dir.
- * Gerçek pozitif oranı 0'dır.
- * Test setinin %33'ü pozitiftir.
- * Pozitif örnekleri sınıflandırmak daha zordur.

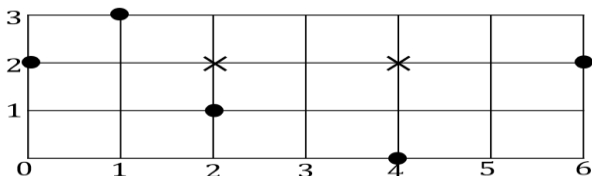
- a.4 b.3 c.2 d.1

12. Kümeleme ile ilgili aşağıda yazılanlardan hangisi yada hangileri doğrudur?

- I. Kümeleme analizi, test örneklerinin hangi kümeye ait olduklarını tahmin etmek için yapılır.
- II. Bir veri setinden elde edilebilecek küme sayısı sabittir
- III. Kümeleme analizi sonucu herhangi bir kümeye dahil olmayan örnekler anomali olabilirler.

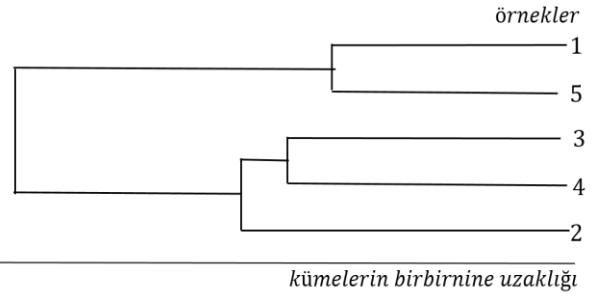
- a. Yalnız III b. I ve III c. I,II,III d. Yalnız I.

13. Aşağıda küme merkezleri \times ile, örnekler noktalarla gösteriliyor. K-ortalama algoritmasının bir iterasyonu sonucunda yeni küme merkezleri hangi koordinatlar olur?



- a. (1,2) ve (5,1) b. (1,1) ve (5,1)
c. (3,2) ve (3,2) d. (1,2) ve (2,3)

14. Aşağıda numaralandırılmış beş örnek ile yapılan hiyerarşik kümelemeye ait dendogram gösterilmektedir.



Aşağıdakilerin hangileri doğrudur?

- I. Birbirine en yakın iki örnek 1 ve 5 dir.
- II. Veri setinde muhtemelen iki küme vardır.
- III. En sondan bir önce oluşan küme, 2'nin 3 ve 4'ün oluşturduğu kümeye eklenmesiyle oluşur.

- a. I ve II b. I ve III c. I,II,III d. Yalnız I.

15 ve 16 soruları aşağıda gösterilen işlemler listesine göre çözünüz.

İşlem No	Ürünler
1	{Kebap, Salata, Çorba, Sütlaç}
2	{Kebap, Ayran, Salata, Döner}
3	{Ayran, Salata, Çorba, Döner}
4	{Kebap, Ayran, Salata, Çorba}
5	{Kebap, Ayran}

15. Salata, Çorba \rightarrow Ayran ve Ayran \rightarrow Salata, Çorba birliktelik kurallarının destek değeri nedir?

- a.0.4 b.0.1 c.0.6 d.0.2

16. Ayran, Döner \rightarrow Salata birliktelik kuralının güven değeri nedir?

- a.0.8 b.1 c.0.5 d.0.66

17. Aşağıdakilerin kaç tanesi doğrudur?

- *Apriori algoritması, sık görülmeyen bir ürünü içeren ürün kümelerinin de sık görülmeyeceği görüşüne dayanır. +
- *Bir kuralın destek değerinin 0.5 olması, yapılan işlemlerin %50 sinde bu kuraldaki ürünlerin beraber alındığı anlamına gelir. +
- * $A \rightarrow B$ şeklindeki bir birliktelik kuralı, A'yı almanın B'yi de almaya neden olacağını anlatır.
- * $A \rightarrow B$ şeklindeki bir birliktelik kuralının güveni 0.7 ise A'yı alanların %70'nin B'yi de aldığı sonucuna varırız. +

- a.4 b.3 c.2 d.1