

-SINAV SÜRENİZ 70 DAKİKADIR.

-CEVAPLARINIZI EKAMPÜS'E YÜKLEMELİK İÇİN +15 DAKİKA DAHA EK SÜRENİZ VARDIR.

-VERİLEN SÜRENİN SONUNDA HERHANGİ BİR NEDENLE CEVAPLARINIZI EKAMPÜS'E YÜKLEYEMEMENİZ DURUMUNDA SINAVA GİRMEMİŞ KABUL EDİLİRSİNİZ. MAIL VEYA BAŞKA BİR KANALDAN GÖNDERİLMEMEYE ÇALIŞILAN CEVAPLAR KABUL EDİLMEYECEKTİR.

-EKAMPÜS DOSYA YÜKLEME LİMİTİ 10 MB'DIR.

-SORULARDAN PUAN ALABİLMENİZ İÇİN ÇÖZÜMLERİNİZDE KULLANDIĞINIZ/TÜRETTİĞİNİZ DENKLEMLERİ AÇIK BİR ŞEKİLDE YAZMANIZ GEREKMEKTEDİR.

Sorular

1) [30p] Yanda iki boyutlu bir uzayda bulunan yedi adet noktanın koordinatları (x,y) verilmiştir. Bu noktaları K-means algoritmasını kullanarak K1 ve K2 gibi iki kümeye ayırmak istiyoruz.

Bildiğiniz üzere K-means algoritması bir dizi adımdan (iterasyondan) oluşmaktadır. Herhangi bir adımı i sembolü ile gösterecek olursak bu adımdaki K1 kümesinin merkezi μ_{1i} ve K2 kümesinin merkezi de μ_{2i} ile ifade edilir.

	x	y
A	3	8
B	-2	4
C	4	6
D	-1	1
E	5	3
F	6	2
G	9	4

Buna göre,

i) K-means algoritmasını bu yedi noktayı kullanarak tam 4 adım boyunca çalıştırıp her adım için bir sonraki sayfadaki küme merkezleri&noktaların merkezlere mesafeleri tablolarını doldurunuz. Tablolarda noktaların merkezlere olan mesafeleri $d(\text{Nokta}, \mu_{1i})$ ve $d(\text{Nokta}, \mu_{2i})$ şeklinde gösterilmiştir. Noktaların merkezlere olan mesafelerini hesaplamak için Mahhattan (City-block) mesafesini kullanmalısınız (başka bir mesafe ölçütü kullanırsanız sorudan puan alamazsınız).

$i = 1$ nolu adım için rasgele seçilen küme merkezleri $\mu_{11} = (4, 6)$ ve $\mu_{21} = (9, 4)$ şeklindedir. Her adım sonunda her bir noktanın hangi kümeye düştüğünü (K1 veya K2) noktanın hemen önüne yazınız (Örneğin, K1 C gibi).

ii) 4. adımın sonunda K1 ve K2 kümeleri için ayrı ayrı küme içi varyansları ($\text{var}(K1)$ ve $\text{var}(K2)$) hesaplayarak rapor ediniz. Küme içi varyansı şu formülle hesaplayabilirsiniz:

$$\text{var}(K) = \frac{1}{N} \sum_{\text{Nokta} \in K} d(\text{Nokta}, \mu_K)$$

Burada, N sembolü K kümesindeki nokta sayısını, μ_K sembolü K kümesinin merkezini, $d()$ fonksiyonu ise K kümesindeki herhangi bir nokta ile bu kümenin merkezi arasındaki Manhattan mesafesini göstermektedir.

$$\mu_{11} = (4, 6)$$

$$\mu_{21} = (9, 4)$$

		d(Nokta, μ_{11})	d(Nokta, μ_{21})
K1	A		
	B		
	C	0	
	D		
	E		
	F		
K2	G		0

$$\mu_{12} = ?$$

$$\mu_{22} = ?$$

		d(Nokta, μ_{12})	d(Nokta, μ_{22})
A			
B			
C			
D			
E			
F			
G			

$$\mu_{13} = ?$$

$$\mu_{23} = ?$$

		d(Nokta, μ_{13})	d(Nokta, μ_{23})
A			
B			
C			
D			
E			
F			
G			

$$\mu_{14} = ?$$

$$\mu_{24} = ?$$

		d(Nokta, μ_{14})	d(Nokta, μ_{24})
A			
B			
C			
D			
E			
F			
G			

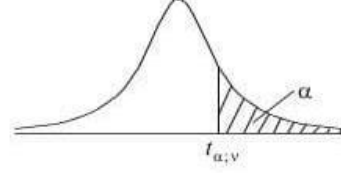
2) [40p] Küresel hizalama (global alignment) yöntemini kullanarak aşağıda verilen hizalama matrisini doldurunuz. Sorudan puan alabilmeniz için her hücreye hangi yönden gelindiğini gösteren yön oklarını da eksiksiz bir şekilde matriste göstermeniz gerekmektedir. Eşleşme (match) skoru +3, eşleşmeme (mismatch) skoru -2, boşluk açma/uzatma (gap open/extend) skoru -1 birim olarak kabul edilecektir. Matrisi doldurduktan sonra backtracking yöntemini kullanarak iki dizilimin nasıl eşleştirildiğini gösteren herhangi bir eşleşmeyi de yazınız.

	-	C	G	T	A	G	G
-							
T							
G							
A							
C							

3) [30p] $X = [9, -4, 7, 6, 1, 6, 5, 8, -2]$ ve $Y = [7, 9, -5, 3, 7, 6, 4]$ gibi iki dağılımımız olsun. Bu dağılımların ortalamaları arasında anlamlı bir farkın olup olmadığını t-skorunu hesaplayarak $p=0,05$ anlamlılık düzeyi için gösteriniz. Sorudan puan alabilmek için türetilmesi gereken denklemleri açık bir şekilde yazmanız gerekmektedir.

Table of the Student's t -distribution

The table gives the values of $t_{\alpha;v}$ where
 $\Pr(T_v > t_{\alpha;v}) = \alpha$, with v degrees of freedom



$\alpha \backslash v$	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.076	31.821	63.657	318.310	636.620
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291