



**SEMESTER 5**  
**PEMBELAJARAN**  
**MESIN**  
**AIK21356**

**DMW++**

---

**DIKLAT HMIF UNDIP**

## DAFTAR ISI

UJIAN TENGAH SEMESTER 2014/2015 .....	3
UJIAN AKHIR SEMESTER 2017/2018 .....	4
UJIAN AKHIR SEMESTER 2018/2019 .....	6
UJIAN AKHIR SEMESTER 2019/2020 .....	9
UJIAN TENGAH SEMESTER 2020/2021 .....	12
UJIAN AKHIR SEMESTER 2020/2021 .....	15

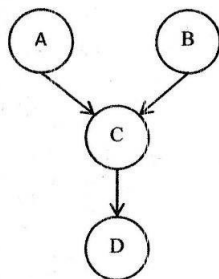
**UJIAN TENGAH SEMESTER 2014/2015****UJIAN TENGAH SEMESTER  
TAHUN AJARAN 2014/2015 SEMESTER GASAL**

Mata Kuliah : Machine Learning  
 Hari/Tanggal : Selasa/4 November 2014  
 Waktu : 120 menit  
 Sifat : Buka Buku (*Open Book*)

Keterangan : Angka pada awal setiap pertanyaan merupakan bobot nilai untuk jawaban dari soal yang bersangkutan

1. (20) Misalkan kita memiliki problem klasifikasi untuk kelas biner yaitu  $C_1$  dan  $C_2$  dan sebuah fitur yang memiliki nilai biner  $x$  dan  $\sim x$ . Diasumsikan *prior probability* dari  $C_2$  adalah  $P(C_2) = 0,41$ . Jika total data sampel yang termasuk kelas  $C_1$  sebanyak 350 dan total data yang termasuk kelas  $C_2$  adalah 670. Berdasarkan data sampel tersebut ternyata diketahui bahwa 292 dari data sampel yang termasuk kelas  $C_1$  memiliki fitur  $x$  sedangkan 512 dari data sampel yang termasuk kelas  $C_2$  memiliki fitur  $\sim x$ . Berdasarkan data tersebut, tentukan probabilitas posterior dari kelas-kelasnya menggunakan *Bayesian Rule*! Jika sebuah data baru memiliki fitur  $\sim x$ , maka termasuk dalam kelas apakah data baru tersebut?

2. (40) Diketahui struktur dari sebuah *Bayesian Networks* sebagai berikut:



Berdasarkan struktur *Bayesian Networks* tersebut, misalnya setiap variable bernilai biner dan *conditional probabilities* dari struktur tersebut adalah:

$$P(A = 1) = 0,80$$

$$P(B = 1) = 0,20$$

$$P(C = 1|A = 0, B = 0) = 0,30$$

$$P(C = 1|A = 0, B = 1) = 0,85$$

$$P(C = 1|A = 1, B = 0) = 0,45$$

$$P(C = 1|A = 1, B = 1) = 0,65$$

$$P(D = 1|C = 1) = 0,75$$

$$P(D = 1|C = 0) = 0,30$$

- Susunlah tabel *joint probability* dari distribusi yang direpresentasikan menggunakan model tersebut!
- Berapakah nilai  $P(A, B, C, D)$  terbesar?
- Hitung nilai *marginal probabilities*  $P(A)$ !

3. (40) Jika diketahui data sebagai berikut:

Data	$F_1$	$F_2$
1	1	1
2	2	2
3	3	3
4	4	4
5	1	2
6	4	3
7	1	3
8	3	4

Terapkan *hierarchical clustering* menggunakan pendekatan *single linkage* (jarak terdekat) dan bangun dendogramnya!

**UJIAN AKHIR SEMESTER 2017/2018**

KEMENTERIAN RISET, TEKNOLOGI DAN PENDIDIKAN TINGGI  
**UNIVERSITAS DIPONEGORO**  
**FAKULTAS SAINS DAN MATEMATIKA**  
Jalan Prof. H. Soedarto, SH. Tembalang Semarang 50275

**UJIAN AKHIR SEMESTER (UAS) GASAL 2017/ 2018**  
**DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA**

Mata Kuliah : Pembelajaran Mesin

Hari, Tanggal : Senin, 4 Desember 2017

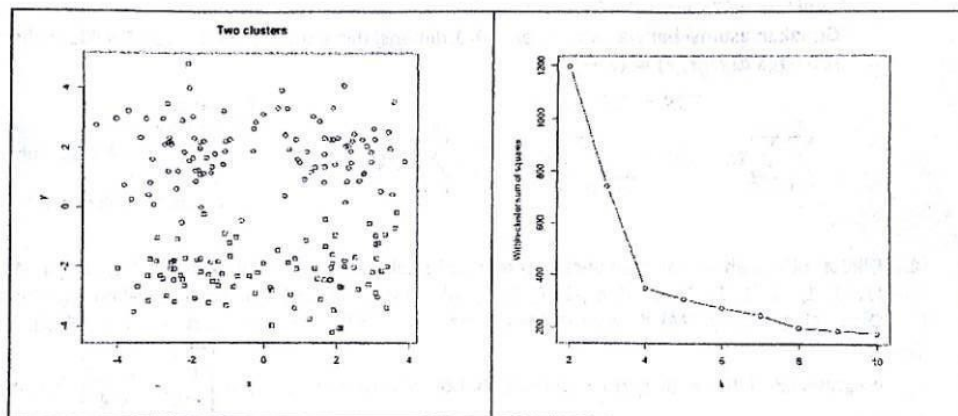
Pengampu : Dr. Retno Kusumaningrum, M.Kom.

Waktu : 13.00 – 14.40 (100 menit)

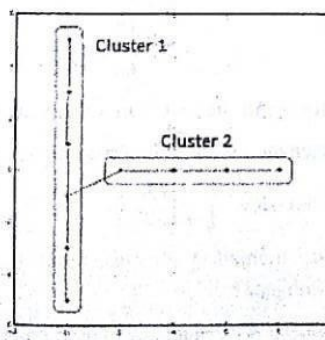
Khadijah, S.Kom., M.Cs.

Sifat Ujian : Open 1 lembar catatan ukuran A5

1. a. Gambar di sebelah kiri menunjukkan hasil clustering sejumlah data menggunakan k-means clustering dengan  $k=2$ , sedangkan grafik di sebelah kanannya menunjukkan plot hasil clustering menggunakan k-means clustering dengan berbagai jumlah cluster ( $k = 2, 3, 4, \dots, 10$ ) dan Sum of Squared Errors (SSE) yang didapat pada setiap  $k$ . Menurut Anda berapa jumlah cluster terbaik pada tersebut? Berikan penjelasan.



- b. Gambar berikut menunjukkan plot 10 titik (data) dan cluster yang terbentuk pada titik-titik tersebut. Jelaskan jawaban Anda dengan jelas apakah algoritma k-means clustering dapat digunakan untuk mendapatkan hasil clustering seperti gambar tersebut? Jika jawaban Anda tidak, jelaskan strategi yang dapat dilakukan dengan algoritma k-means untuk mendapatkan hasil clustering yang mendekati gambar tersebut.





2. Diketahui 5 buah objek yang akan dikluster menggunakan algoritma *agglomerative hierarchical clustering (AHC)*. Dengan menggunakan similarity matrix seperti gambar berikut, lakukanlah clustering dan gambarkan dendogramnya menggunakan complete link AHC.  
(Perhatikan bahwa nilai yang di matriks adalah similarity, bukan distance)

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

3. Ukuran untuk evaluasi clustering secara unsupervised adalah separation antar cluster dan cohesion dalam setiap cluster.

- a. Jelaskan apa yang dimaksud dengan ukuran cohesion dan separation tersebut?
- b. Buktikan bahwa jumlah Sum of Squared Errors - SSE (menunjukkan ukuran cohesion) dan Between Group Sum of Squares - SSB (menunjukkan ukuran separation) adalah konstan, yaitu Total Sum of Squares (TSS).  
Gunakan asumsi bahwa datanya adalah 1 dimensi dan ukuran jarak berupa Euclidean distance, sehingga  $dist(x, y) = (x - y)^2$

$$TSS = SSE + SSB$$

$$\sum_{i=1}^K \sum_{x \in C_i} (x - c)^2 = \sum_{i=1}^K \sum_{x \in C_i} (x - c_i)^2 + \sum_{i=1}^K |C_i| (c - c_i)^2$$

Keterangan:

$K$  : jumlah cluster

$c$  : centroid untuk keseluruhan data

$C_i$  : cluster ke- $i$

$c_i$  : centroid untuk cluster ke- $i$

4. Diketahui 5 buah vektor akan di-cluster menjadi 2 kelompok seperti berikut:  $(-1, 1, 1, -1)$ ,  $(-1, -1, 1, 1)$ ,  $(1, -1, 1, 1)$ ,  $(1, 1, -1, -1)$ , dan  $(-1, 1, -1, 1)$ . Jika clustering dilakukan menggunakan algoritma Self Organizing Map (SOM) dengan inisialisasi parameter berikut: learning rate ( $\alpha$ ) = 0,5 dan radius

neighborhood ( $R$ ) = 0. serta inisialisasi bobot seperti berikut:  $\mathbf{w} = \begin{bmatrix} -0,5 & 0,2 \\ 0,3 & -0,7 \\ -0,7 & 0,4 \\ 0,3 & -0,5 \end{bmatrix}$ . Gambarkan

arsitektur jaringannya dan hitunglah bobot hasil revisi pada iterasi pertama untuk data yang pertama.

5. Diketahui 6 buah data dengan 2 atribut yang direpresentasikan dalam matriks A, akan dilakukan reduksi dimensi menggunakan algoritma *principal component analysis (PCA)*.

$$\mathbf{A} = \begin{bmatrix} 8 & 6 \\ 9 & 3 \\ 4 & 7 \\ 5 & 2 \\ 7 & 10 \\ 9 & 8 \end{bmatrix}$$

Eigenvalue dan eigenvector yang didapat dari matriks kovarian dari data tersebut adalah : 1)

Eigenvalue = 9,274 dan eigenvector yang bersesuaian adalah  $\begin{bmatrix} 0,1222 \\ 0,9925 \end{bmatrix}$ ; 2) Eigenvalue = 4,326 dan

eigenvector yang bersesuaian adalah  $\begin{bmatrix} -0,9925 \\ 0,1222 \end{bmatrix}$ .

- a. Jika data akan direduksi menjadi 1 dimensi, manakah eigenvector yang Anda pilih sebagai *principal component*? Mengapa?
- b. Reduksilah data pada matriks A menjadi 1 dimensi menggunakan *principal component* yang telah dipilih pada langkah a. (Representasikan dalam matriks ukuran  $6 \times 1$ ).

# UJIAN AKHIR SEMESTER 2017/2018



KEMENTERIAN RISET, TEKNOLOGI DAN PENDIDIKAN TINGGI  
UNIVERSITAS DIPONEGORO  
FAKULTAS SAINS DAN MATEMATIKA  
Jalan Prof. H. Soedarto, SH. Tembalang Semarang 50275

UJIAN AKHIR SEMESTER (UAS) GASAL 2018/ 2019  
DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA

Mata Kuliah : Pembelajaran Mesin

Hari, Tanggal : Jumat, 7 Desember 2018

Pengampu : Dr. Retno Kusumaningrum, M.Kom.

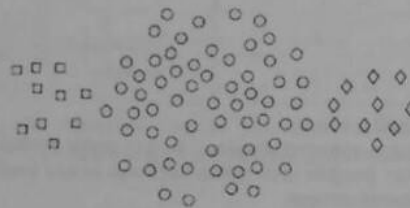
Waktu : 08.00 – 09.40 (100 menit)

Khadijah, S.Kom., M.Cs.

Sifat Ujian : Close Book

1. Clustering dan K-Means

- [Skor 10] Jelaskan perbedaan tujuan antara *clustering for understanding* dan *clustering for utility*. Berikan masing-masing 1 (satu) buah contoh.
- [Skor 10] Jelaskan pendapat Anda apakah k-means clustering dapat digunakan untuk mengkluster data berikut? (simbol yang berbeda menunjukkan cluster data)



- [Skor 20] Diketahui 5 buah objek yang akan dikluster menggunakan algoritma *agglomerative hierarchical clustering (AHC)*. Dengan menggunakan similarity matrix seperti gambar berikut, lakukanlah clustering dan gambarkan dendogramnya menggunakan complete link AHC. (Perhatikan bahwa nilai yang di matriks adalah similarity, bukan distance)

	p1	p2	p3	p4
p1	1.00	0.10	0.41	0.55
p2	0.10	1.00	0.64	0.47
p3	0.41	0.64	1.00	0.44
p4	0.55	0.47	0.44	1.00

3. SOM (Self Organizing Map)

- [Skor 5] Jelaskan apa yang dimaksud dengan SOM merupakan jaringan syaraf tiruan yang berbasis kompetisi.
- [Skor 5] Diketahui 5 buah vektor akan di-cluster menjadi 2 kelompok seperti tabel berikut. Jika clustering dilakukan menggunakan algoritma SOM dengan inisialisasi parameter learning rate ( $\alpha$ ) = 0,5 dan radius neighborhood ( $R$ ) = 0, serta inisialisasi bobot seperti berikut:

Data ke-	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>
1	Yes	No	Yes	Yes
2	Yes	Yes	No	No
3	No	Yes	Yes	No
4	No	Yes	Yes	Yes
5	Yes	No	No	No

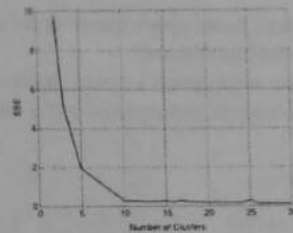
$$w = \begin{bmatrix} 0,5 & 0,9 \\ 0,7 & 0,6 \\ 0,4 & 0,7 \\ 0,2 & 0,3 \end{bmatrix}$$

Pada iterasi pertama untuk data pertama, bobot-bobot menuju neuron manakah yang akan diupdate?



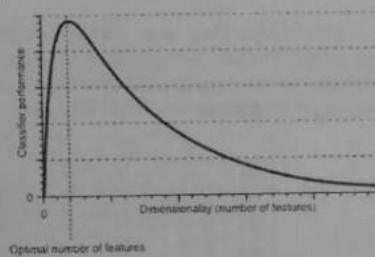
Scanned with  
CamScanner

- c. [Skor 5] Berapakah bobot revisi untuk neuron output (yang dipilih untuk update pada soal b) yang berasal dari node input pertama.
4. [Skor 10] Cluster evaluation (skor masing-masing= 2,5)
- [True/ False] SSE (Sum of Squared Error) merupakan salah satu ukuran cohesion.
  - [True/ False] Nilai separation dan cohesion hasil clustering berbanding lurus. Artinya, jika nilai separation tinggi, maka nilai cohesion juga akan tinggi.
  - [True/ False] Saat menggunakan proximity measure berupa jarak (distance), maka semakin rendah nilai silhouette coefficients menunjukkan hasil clustering yang semakin baik.
  - Grafik berikut menunjukkan hubungan antara jumlah cluster dengan nilai SSE yang didapat. Berdasarkan grafik tersebut berapakah jumlah cluster yang optimal?  $\zeta$



5. Curse of dimensionality:

- a. [Skor 10] Jelaskan apa yang dimaksud dengan grafik berikut. Mengapa hal seperti itu dapat terjadi? Jelaskan jawaban anda dalam kaitannya antara overfitting, jumlah fitur, jumlah data sample, dan sparsity sample.



- b. [Skor 10] Diketahui 6 buah data dengan 2 atribut yang direpresentasikan dalam tabel berikut, akan dilakukan reduksi dimensi menggunakan algoritma *principal component analysis* (PCA).

Data ke-	$x_i$	$y_i$	$x_i - \bar{x}$	$y_i - \bar{y}$
1	8	6	1	0
2	9	3	2	-3
3	4	7	-3	1
4	5	2	-2	-4
5	7	10	0	4
6	9	8	2	2
Mean	7	6		

Eigenvalue dan eigenvector yang didapat dari matriks kovarian dari data tersebut adalah :

- 1) Eigenvalue = 9,274 dan eigenvector yang bersesuaian adalah  $\begin{bmatrix} 0,1222 \\ 0,9925 \end{bmatrix}$
- 2) Eigenvalue = 4,326 dan eigenvector yang bersesuaian adalah  $\begin{bmatrix} -0,9925 \\ 0,1222 \end{bmatrix}$

Reduksilah data tersebut menjadi 1 dimensi menggunakan algoritma PCA.



Scanned with  
CamScanner

## 6. Ensemble Learning

- a. [Skor 2,5] [True/ False] Boosting merupakan algoritma ensemble learning dimana proses pelatihan pada weak learners yang digunakan, dapat dilakukan secara paralel.
- b. [Skor 2,5] Klasifikasi kanker payudara (positif atau negatif) dapat dilakukan berdasarkan beberapa jenis inputan: i) citra mammography; ii) pola ekspresi gen; iii) data klinis pasien. Jika ingin dibangun ensemble classifier menggunakan ketiga jenis inputan yang berbeda tersebut. Maka termasuk kategori manakah alasan dilakukan ensemble learning pada contoh tersebut (pilih salah satu):
- A. Data selection  
B. Data fusion  
C. Divide and conquer  
D. Too much or too little data
- c. [Skor 10] Menggunakan contoh kasus pada soal b. Classifier yang digunakan adalah k-NN (k=5) untuk masing-masing jenis inputan tersebut. Jika output dari k-NN untuk masing-masing jenis inputan adalah sebagai tabel berikut.

Ensemble classifier (berdasarkan jenis inputan)	Class label dari 5 tetangga terdekat yang dipilih oleh k-NN				
	1	2	3	4	5
1 (citra mammography)	+	-	+	-	-
2 (data ekspresi gen)	+	-	-	-	-
3 (data klinis pasien)	-	+	+	-	+

Maka output akhir dari ensemble classifier jika digunakan Algebraic combiners dengan maximum rule adalah .... (tuliskan langkah-langkah penyelesaiannya).



Scanned with  
CamScanner





KEMENTERIAN RISET, TEKNOLOGI, DAN PENDIDIKAN TINGGI  
UNIVERSITAS DIPONEGORO  
FAKULTAS SAINS DAN MATEMATIKA

Jalan Prof. H. Soedarto, SH. Tembalang Semarang 50275; Telp : (024) 7474754; Fax : (024) 76480690

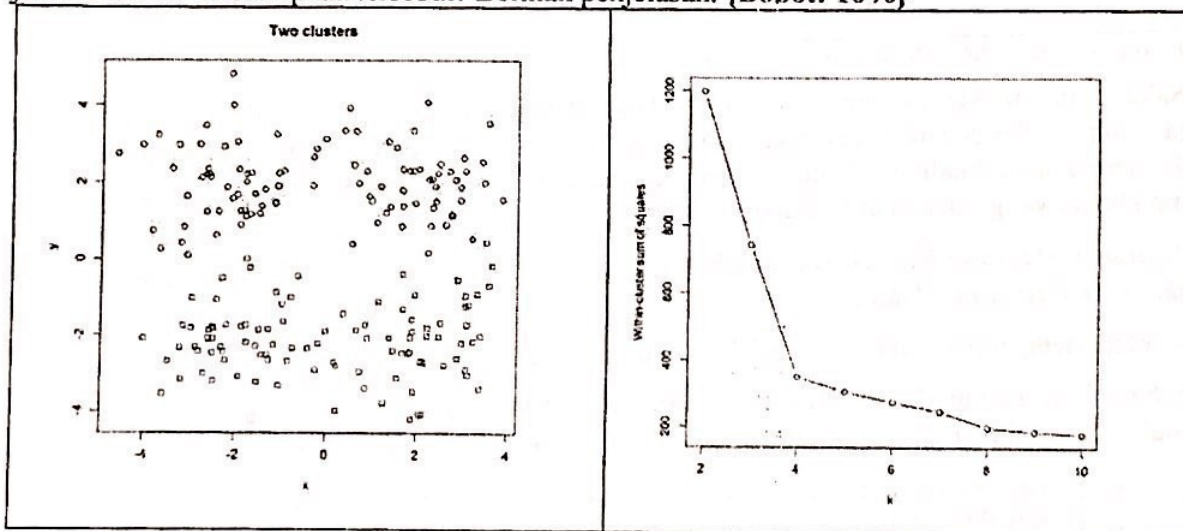
UJIAN AKHIR SEMESTER GASAL 2019/2020

Mata Kuliah	:	Pembelajaran Mesin
Kelas	:	A/B/C
Pengampu	:	Dr. Retno Kusumaningrum, S.Si, M.Kom. Rismiyati, B.Eng, M.Cs
Departemen	:	Informatika
Program Studi	:	S1 Informatika
Hari / Tanggal	:	Jumat, 6 Desember 2019
Jam / Ruang	:	08.00 – 09.40 WIB (100 menit) / A203, A204, I205
Sifat Ujian	:	Buku tertutup

**Petunjuk Pengerjaan :**

Jawablah soal-soal berikut pada lembar jawab!

1. (a) Tulislah algoritma untuk melakukan clustering dengan menggunakan K-Means Clustering. Jelaskan cara kerja algoritma yang anda tulis tersebut! {10%}
- (b) Gambar di sebelah kiri menunjukkan hasil clustering sejumlah data menggunakan k-means clustering dengan  $k=2$ , sedangkan grafik di sebelah kanannya menunjukkan plot hasil clustering menggunakan k-means clustering dengan berbagai jumlah cluster ( $k = 2, 3, 4, \dots, 10$ ) dan Sum of Squared Errors (SSE) yang didapat pada setiap  $k$ . Menurut Anda berapa jumlah cluster terbaik pada tersebut? Berikan penjelasan. {Bobot: 10%}

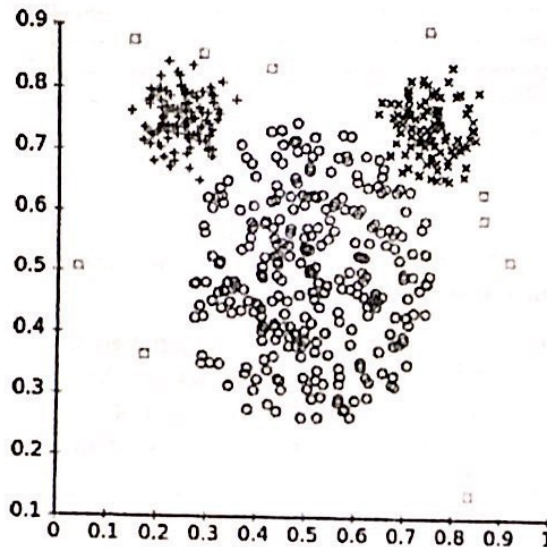


2. {bobot: 20%} Diketahui 5 buah objek yang akan dikluster menggunakan algoritma *agglomerative hierrarchical clustering* (AHC). Dengan menggunakan matriks jarak seperti gambar berikut, lakukanlah clustering dan gambarkan dendogramnya menggunakan complete link AHC. {Bobot 20%}

	1	2	3	4	5
1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	11	10	2	8	0

3. (a) Untuk Data pada gambar berikut, manakah yang akan memberikan hasil cluster yang lebih baik antara K-means dan EM Algorithm. Jelaskan pilihan anda dengan menitikberatkan pada perbedaan antar kedua algoritma tersebut! {Bobot 10%}

Original Data



(b) Salah satu metode evaluasi clustering adalah dengan unsupervised method, yaitu dengan kohesi dan separasi. Jelaskan bagaimana kohesi dan separasi dapat digunakan untuk menentukan kualitas klustering secara keseluruhan, maupun kualitas masing-masing kluster yang dihasilkan! {Bobot: 10%}

4. (a). "Dalam melakukan klasifikasi atau clustering, semakin banyak fitur yang digunakan, semakin baik hasil yang didapat."

Apakah anda setuju dengan pernyataan di atas! Jelaskan argumentasi anda! {bobot: 10%}

(b) Diketahui 6 buah data dengan 2 atribut yang direpresentasikan dalam matriks A, akan dilakukan reduksi dimensi menggunakan algoritma *principal component analysis* (PCA).

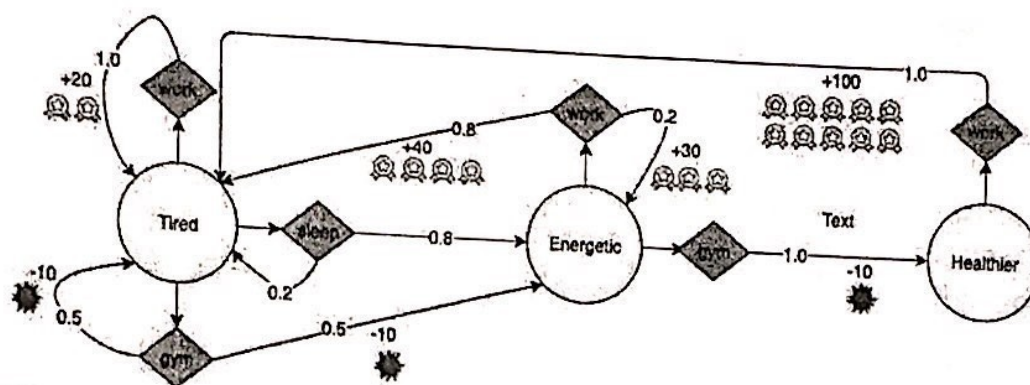
$$A = \begin{bmatrix} 8 & 6 \\ 9 & 3 \\ 4 & 7 \\ 5 & 2 \\ 7 & 10 \\ 9 & 8 \end{bmatrix}$$

Eigenvalue dan eigenvector yang didapat dari matriks kovarian dari data tersebut adalah : 1) Eigenvalue = 4,326 dan eigenvector yang bersesuaian adalah  $\begin{bmatrix} -0.9925 \\ 0.1222 \end{bmatrix}$ , dan 2) Eigenvalue = 9,274 dan eigenvector yang bersesuaian adalah  $\begin{bmatrix} 0.1222 \\ 0.9925 \end{bmatrix}$ .

Reduksilah data pada matriks A menjadi 1 dimensi menggunakan *principal component Analysis*.

(Representasikan dalam matriks ukuran 6 x 1). {Bobot: 15%}

5. Gambar berikut adalah contoh Markov Decision Process. Diketahui seseorang berada pada kondisi Tired. Berdasar gambar tersebut, buatlah policy (gabungan state dan action), yang terdiri dari 4 action yang mampu memaksimalkan keuntungan yang di dapat. Diasumsikan Discounted factor adalah 1. {Bobot: 15%}



Selamat mengerjakan dan semoga sukses.

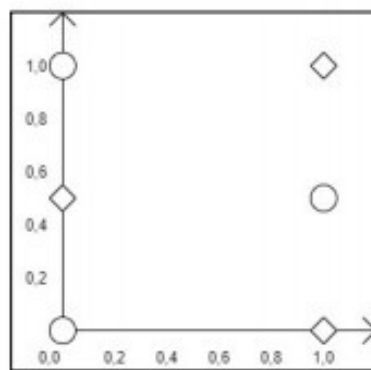


**UJIAN TENGAH SEMESTER 2020/2021****UJIAN TENGAH SEMESTER GASAL 2020/2021**

Mata Kuliah	:	Pembelajaran Mesin
Kelas	:	A, B
Pengampu	:	Dr. Retno Kusumaningrum, S.Si., M.Kom. & Rismiyati, B.Eng., M.Cs.
Departemen	:	Ilmu Komputer/Informatika
Program Studi	:	Informatika
Hari / Tanggal	:	Jumat /15 Oktober 2019
Jam / Ruang	:	10.00 - 11. 40 WIB (100 menit) / <i>Online</i>
Sifat Ujian	:	<i>Open Book</i>

**TIPE 1: JAWABAN SINGKAT**

1. (Nilai 10) Diketahui *training data* yang merepresentasikan 2 buah kelas data, yakni kelas *circle* dan *diamond*, serta dua buah fitur, yakni  $x$  dan  $y$  sebagaimana ditunjukkan pada gambar berikut ini.



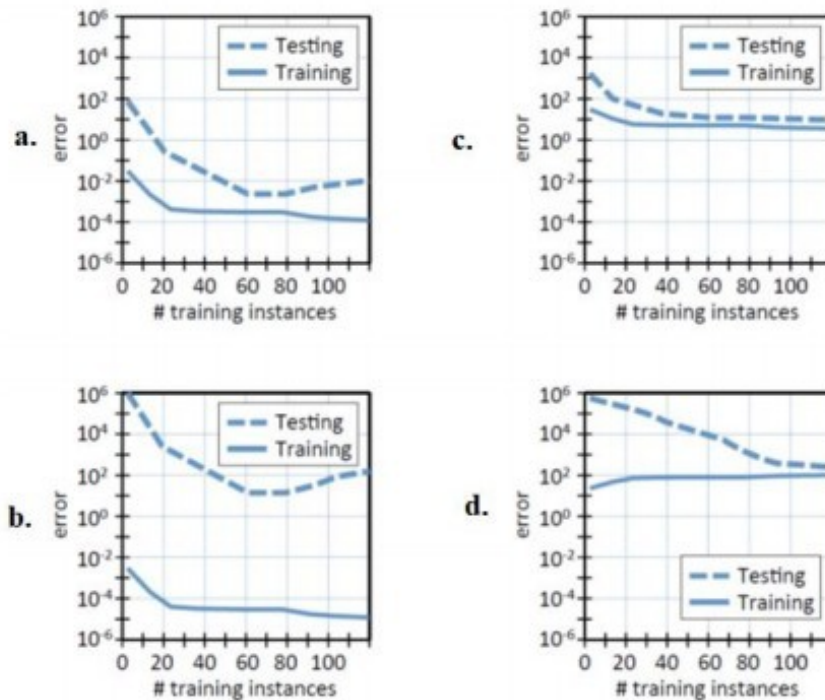
Berdasarkan *training data* tersebut tentukan apakah metode-metode *supervised learning* berikut dapat digunakan untuk menyelesaikan permasalahan klasifikasi tersebut atau tidak.

- 3-Nearest Neighbor : [Ya / Tidak]
  - Decision Tree : [Ya / Tidak]
  - Linear SVM : [Ya / Tidak]
  - SVM dengan Kernel RBF : [Ya / Tidak]
2. (Nilai 10) Setiap grafik berikut menunjukkan nilai *error* yang muncul pada *training data* dan *testing data* untuk sebuah model klasifikasi.

Tentukan apakah masing-masing grafik tersebut menunjukkan salah satu kondisi berikut ini:

- High Bias / High Variance, atau
- High Bias / Low Variance, atau
- Good Balance, atau
- Low Bias / High Variance
- Low Bias / Low Variance





## TIPE 2: URAIAN

3. (Nilai 20) Data penelitian menunjukkan 800 *extrasolar planet* teridentifikasi dalam galaxy kita. Data tersebut menunjukkan apakah planet-planet tersebut dapat ditinggali atau tidak dapat ditinggali, berdasarkan 2 buah fitur yang diamati, yakni ukuran (besar atau kecil) dan orbit (jauh atau dekat). Tabel berikut ini menunjukkan kemungkinan kombinasi fitur serta kondisi apakah planet tersebut dapat ditinggali atau tidak dan jumlah sampel datanya.

Ukuran	Orbit	Ditinggali	Sampel Data
Besar	Jauh	Ya	170
Besar	Dekat	Ya	20
Kecil	Jauh	Ya	45
Kecil	Dekat	Ya	139
Besar	Jauh	Tidak	30
Besar	Dekat	Tidak	130
Kecil	Jauh	Tidak	255
Kecil	Dekat	Tidak	11

Selesaikan permasalahan tersebut menggunakan Decision Tree (ID3).

- Tentukan bagaimana proses splitting dilakukan untuk *training data* tersebut
  - Gambarkan struktur tree yang terbentuk berdasarkan proses splitting sebelumnya.
4. (Nilai 20) Diketahui *training data* yang merepresentasikan 4 buah fitur, yakni demam (tidak, sedang, atau tinggi), mual-muntah (ya atau tidak), diare (ya atau tidak), serta gemetar (ya atau tidak). Adapun kelasnya menunjukkan kondisi seseorang, yakni Sehat (S), Influenza (I), Tifus (T), dan Radang Usus (R).

<b>Id Data</b>	<b>Demam</b>	<b>Mual-Muntah</b>	<b>Diare</b>	<b>Gemetar</b>	<b>Kondisi</b>
D1	Tidak	Tidak	Tidak	Tidak	Sehat
D2	Sedang	Tidak	Tidak	Tidak	Influenza
D3	Tinggi	Tidak	Tidak	Ya	Influenza
D4	Tinggi	Ya	Ya	Tidak	Tifus
D5	Sedang	Tidak	Ya	Tidak	Tifus
D6	Tidak	Ya	Ya	Tidak	Radang Usus
D7	Sedang	Ya	Ya	Tidak	Radang Usus

Jika diketahui sebuah data baru berupa gejala seseorang mengalami mual-muntah dan gemetaran, maka tentukan kelas prediksi dari data baru tersebut menggunakan

- K-Nearest Neighbor*
- Naïve Bayes*

5. **(Nilai 20)** Jika diketahui data training SVM terdiri dari 4 titik sebagai berikut:

$x_1$	$x_2$	Class
1	4	-1
2	3	-1
4	5	+1
5	6	+1

- Tentukan hyperplane dari proses klasifikasi SVM tersebut!
- Gambarkan hyperplane yang sudah diperoleh dan berilah tanda lingkaran pada titik-titik yang merepresentasikan support vector!

6. **(Nilai 20)** Diketahui sebuah data hasil klasifikasi *binary class* untuk deteksi citra kanker payudara meliputi positif terdeteksi kanker dan negatif kanker adalah sebagai berikut:

No	Data	Actual Class	Predicted Class
1	Data 1	NEGATIF	NEGATIF
2	Data 2	NEGATIF	NEGATIF
3	Data 3	NEGATIF	NEGATIF
4	Data 4	NEGATIF	POSITIF
5	Data 5	NEGATIF	NEGATIF
6	Data 6	NEGATIF	NEGATIF
7	Data 7	POSITIF	NEGATIF
8	Data 8	POSITIF	POSITIF
9	Data 9	POSITIF	NEGATIF
10	Data 10	POSITIF	POSITIF

- Buatlah *confusion matrix* untuk hasil klasifikasi tersebut!
- Hitung nilai *accuracy* dan *error rate* dari hasil klasifikasi tersebut!
- Hitung nilai *sensitivity* dan *specificity* dari hasil klasifikasi tersebut!
- Bagaimana kinerja dari algoritma klasifikasi tersebut? Jelaskan!

---- akhir dokumen ----

**UJIAN AKHIR SEMESTER 2020/2021****UJIAN AKHIR SEMESTER GASAL 2020/2021**

Mata Kuliah	:	Pembelajaran Mesin
Kelas	:	A/B/C
Pengampu	:	Dr. Retno Kusumaningrum, S.Si, M.Kom. Rismiyati, B.Eng, M.Cs
Departemen	:	Informatika
Program Studi	:	S1 Informatika
Hari / Tanggal	:	Rabu, 16 Desember 2020
Jam / Ruang	:	13.00-15.00 (sudah terupload)
Sifat Ujian	:	Buku tertutup

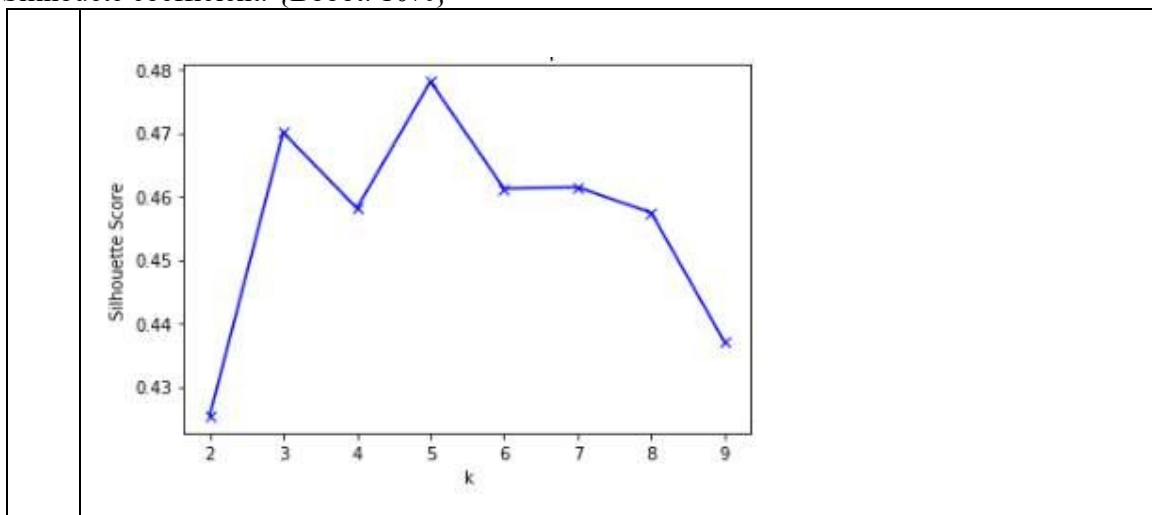
Petunjuk Pengerjaan :

Jawablah soal-soal berikut pada lembar jawab!

1. (a) Ketika menjalankan K means clustering 2 kali berturut-turut dengan jumlah K yang sama, maka hasil cluster yg di dapat akan sama!

Setujukah anda dengan pernyataan tersebut? jelaskan {10%}

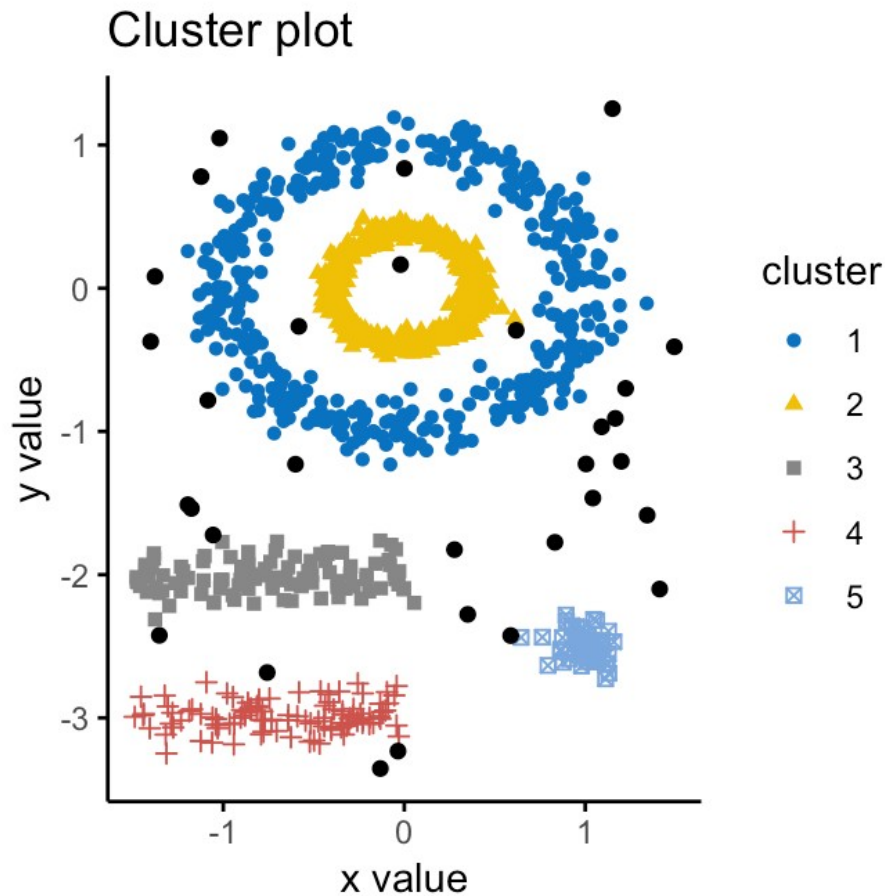
- (b) Gambar berikut kanannya menunjukkan plot Silhouette Coefficient ketika melakukan clustering yang didapat pada setiap k. Dari grafik tersebut, Menurut Anda berapa jumlah cluster terbaik pada data tersebut? Berikan penjelasan dengan menjelaskan definisi dari Silhouette coefficient. {Bobot: 10%}



2. {bobot: 20%} Diketahui 5 buah objek yang akan dikluster menggunakan algoritma agglomerative hierrarchical clustering (AHC). Dengan menggunakan matriks jarak seperti gambar berikut, lakukanlah clustering dan gambarkan dendogramnya menggunakan complete link AHC. {Bobot 20%}

	1	2	3	4	5
1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	11	10	2	8	0

3. (a) Untuk Data pada gambar berikut, Apakah K-means mampu memberikan hasil cluster sebagaimana ditunjukkan pada gambar. Beri penjelasan! {Bobot 10%}



(b) Salah satu metode evaluasi clustering adalah dengan unsupervised method, yaitu dengan kohesi dan separasi. Jelaskan bagaimana kohesi dan separasi dapat digunakan untuk menentukan kualitas klustering secara keseluruhan, maupun kualitas masing-masing kluster yang dihasilkan secara individu! {Bobot: 10%}

4. (a). “Dalam melakukan klasifikasi atau clustering, semakin banyak fitur yang digunakan, semakin baik hasil yang didapat. “

Apakah anda setuju dengan pernyataan di atas! Jelaskan argumentasi anda! {bobot: 10}

(b) Diketahui 6 buah data dengan 2 atribut yang direpresentasikan dalam matriks A, akan dilakukan reduksi dimensi menggunakan algoritma principal component analysis (PCA).

$$A = \begin{bmatrix} 8 & 6 \\ 9 & 3 \\ 4 & 7 \\ 5 & 2 \\ 7 & 10 \\ 9 & 8 \end{bmatrix}$$



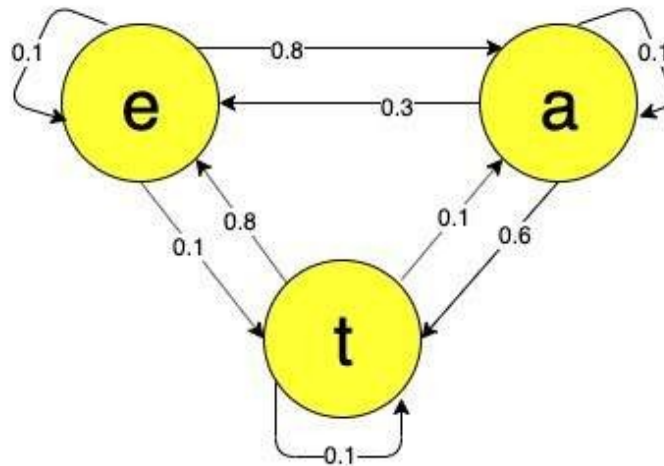
Eigenvalue dan eigenvector yang didapat dari matriks kovarian dari data tersebut adalah : 1)

Eigenvalue = 9,274 dan eigenvector yang bersesuaian adalah  $\begin{bmatrix} 0,1222 \\ 0,9925 \end{bmatrix}$ ; 2) Eigenvalue = 4,326 dan eigenvector yang bersesuaian adalah  $\begin{bmatrix} -0,9925 \\ 0,1222 \end{bmatrix}$ .

Reduksilah data pada matriks A menjadi 1 dimensi menggunakan principal component Analysis.

(Representasikan dalam matriks ukuran 6 x 1). {Bobot: 20%}

5. Dengan menggunakan markov chain



Jika probabilitas huruf pertama =e adalah 50 %, berapakah probabilitas terbentuknya kata eta) ?{Bobot: 10%}

---

Selamat mengerjakan dan semoga sukses.