

DIKLAT
HMIF

2018

PEMBELAJARAN MESIN

AIK21356

SEMESTER 5



DAFTAR ISI

UJIAN TENGAH SEMESTER 2014/2015	1
UJIAN AKHIR SEMESTER 2017/2018.....	2

UJIAN TENGAH SEMESTER 2014/2015

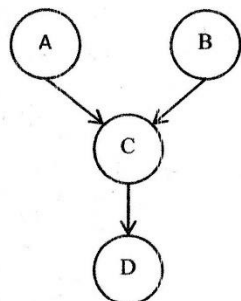
UJIAN TENGAH SEMESTER TAHUN AJARAN 2014/2015 SEMESTER GASAL

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

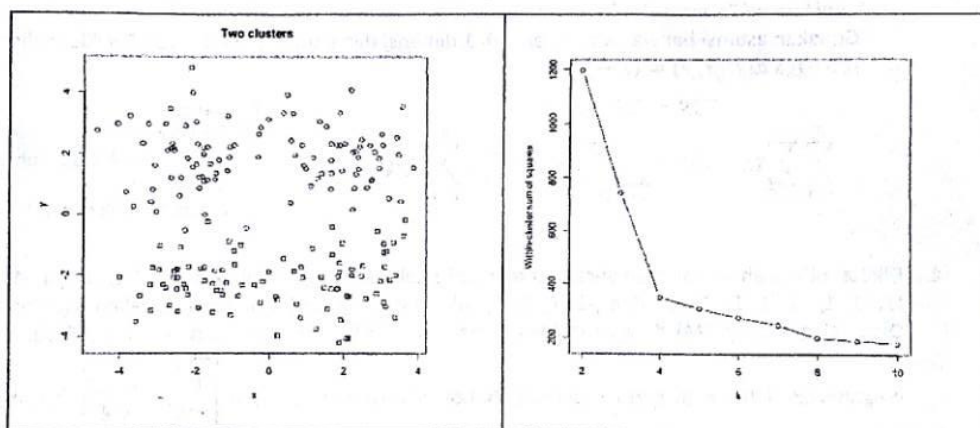
Pengampu : Dr. Retno Kusumaningrum, M.Kom.

Waktu : 13.00 – 14.40 (100 menit)

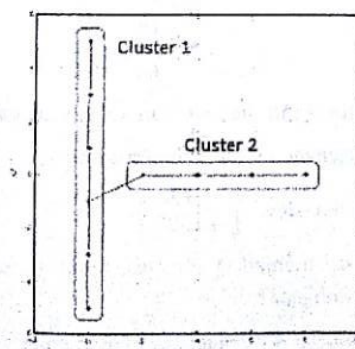
Khadijah, S.Kom., M.Cs.

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 hierrarchical 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	0.41	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 (α) = 0,5 dan radius

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

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

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 x 1).