

Machine Learning

Nearest Neighbors

Ali Ridho Barakbah

Knowledge Engineering Research Group

Soft Computing Laboratory

Department of Information and Computer Engineering

Politeknik Elektronika Negeri Surabaya



Politeknik Elektronika Negeri Surabaya
Departemen Teknik Informatika dan Komputer

Konten

- Nearest Neighbor
- Algoritma 1-NN
- Algoritma k-NN

Tujuan Instruksi Umum

Mahasiswa mampu menyelesaikan masalah – masalah menggunakan metode mesin pembelajaran yang tepat berdasarkan supervised, unsupervised dan reinforcement learning, baik secara individu maupun berkelompok/kerjasama tim.

Tujuan Instruksi Khusus

- Memahami klasifikasi menggunakan k-NN
- Mampu menerapkan k-NN

Nearest Neighbor (NN)

- Merupakan suatu method untuk mengklasifikasikan suatu data baru berdasarkan similaritas dengan labeled data
- Similaritas biasanya memakai metrik jarak
- Satuan jarak umumnya menggunakan euclidian

Nama lain dari NN

- lazy algorithm
- memory-based
- instance-based
- exemplar-based
- case-based
- experience-based

Jenis NN

- 1-NN
 - Pengklasifikasikan dilakukan terhadap 1 labeled data terdekat
- k-NN
 - Pengklasifikasikan dilakukan terhadap k labeled data terdekat
 - $k > 1$

Algoritma 1-NN


- Hitung jarak antara data baru ke setiap labeled data
- Tentukan 1 labeled data yang mempunyai jarak yang paling minimal
- Klasifikasikan data baru ke dalam labeled data tersebut

Contoh kasus 2:

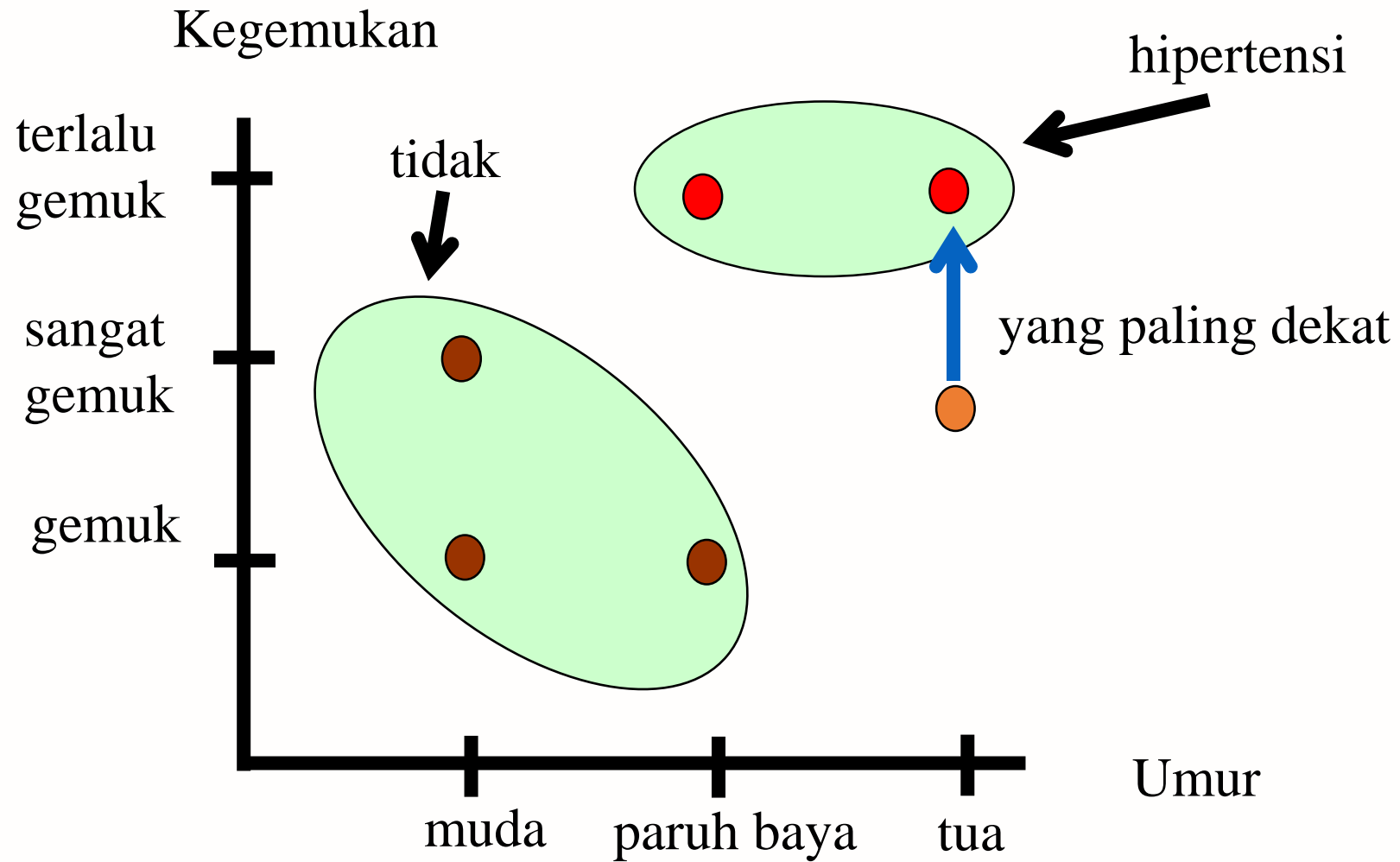
Pengenalan untuk menentukan seseorang itu mempunyai hipertensi atau tidak

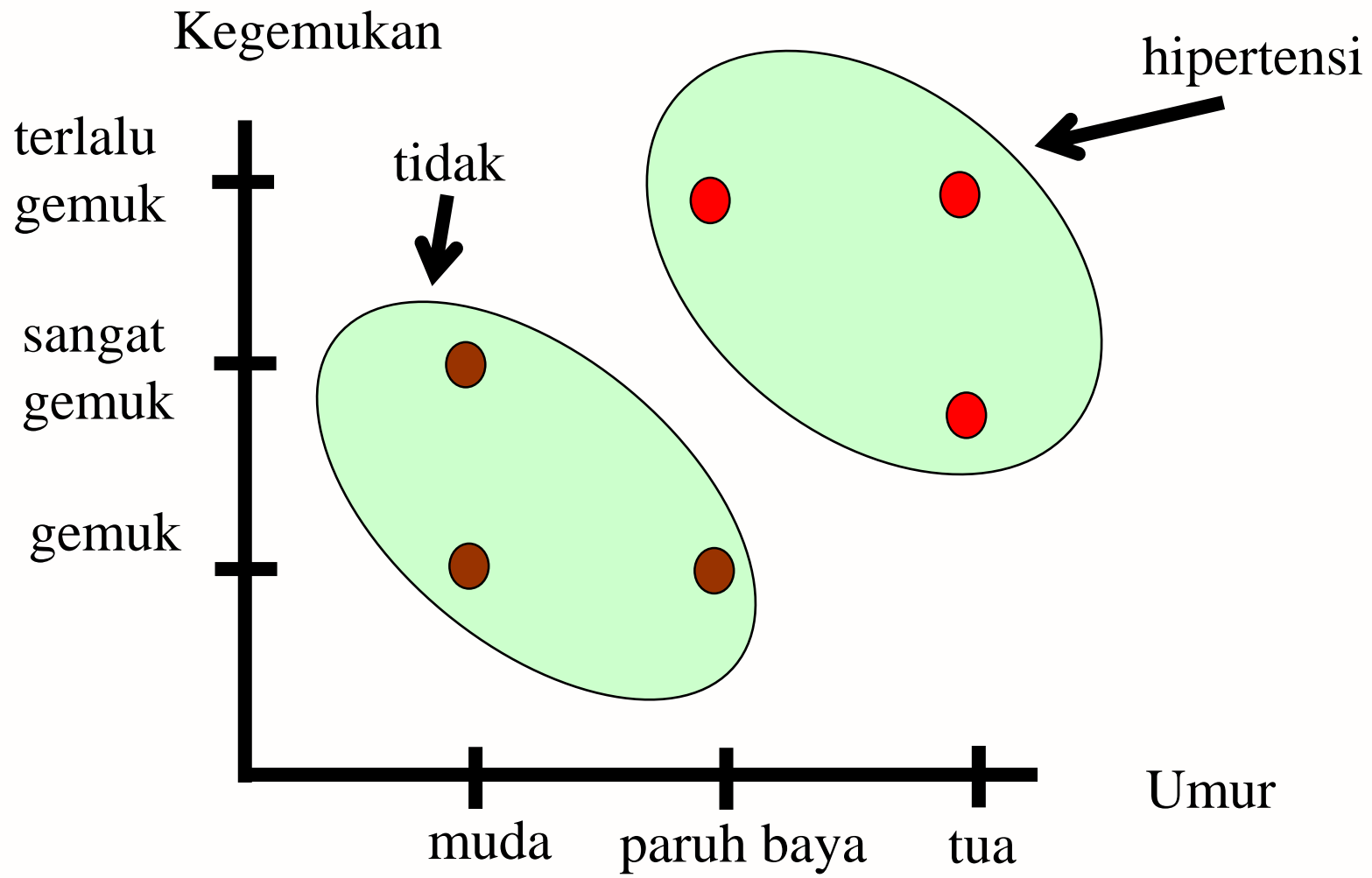
Umur	Kegemukan	Hipertensi
muda	gemuk	Tidak
muda	sangat gemuk	Tidak
paruh baya	gemuk	Tidak
paruh baya	terlalu gemuk	Ya
tua	terlalu gemuk	Ya
tua	sangat gemuk	?

data baru



Penyelesaian dengan 1-NN

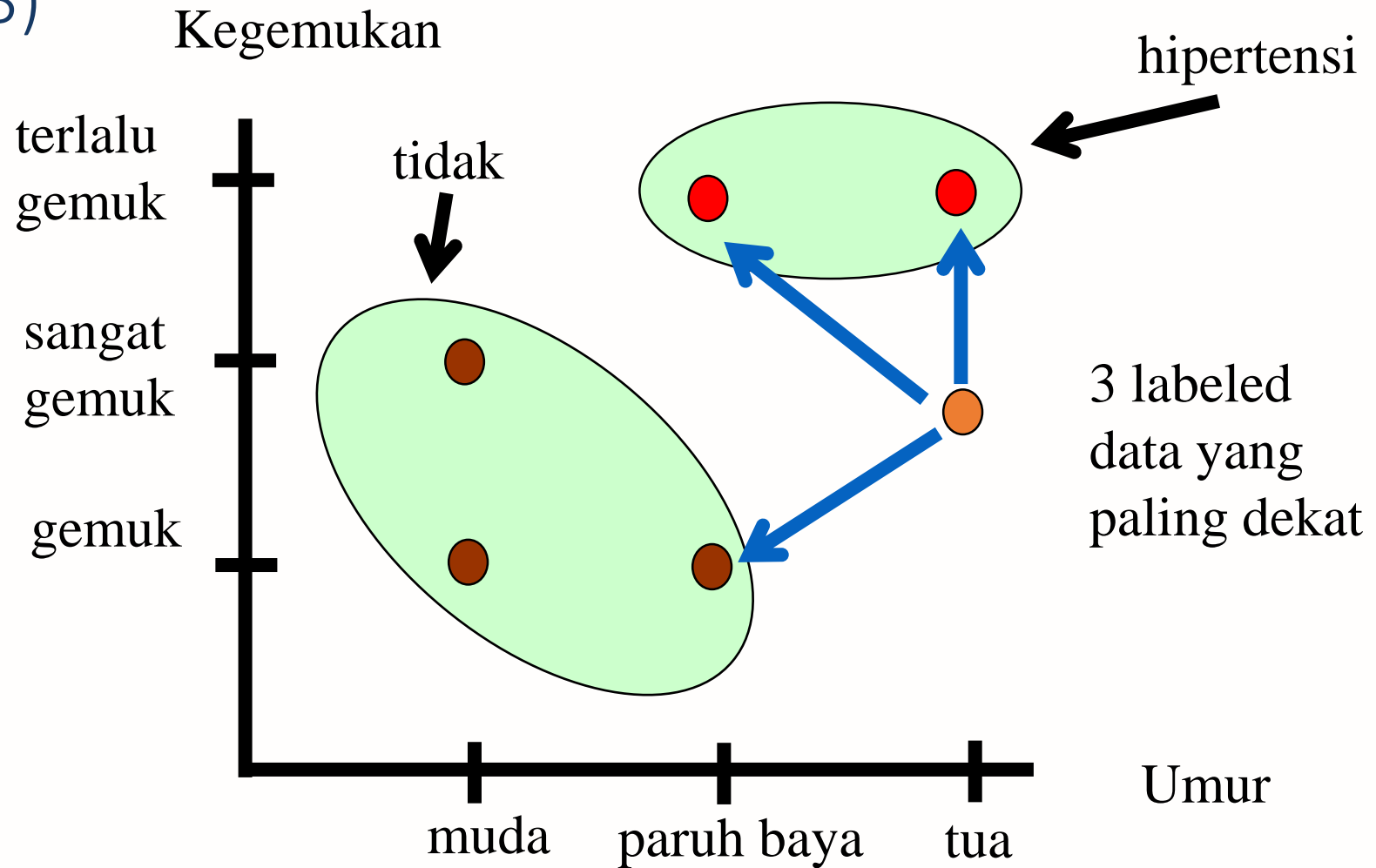


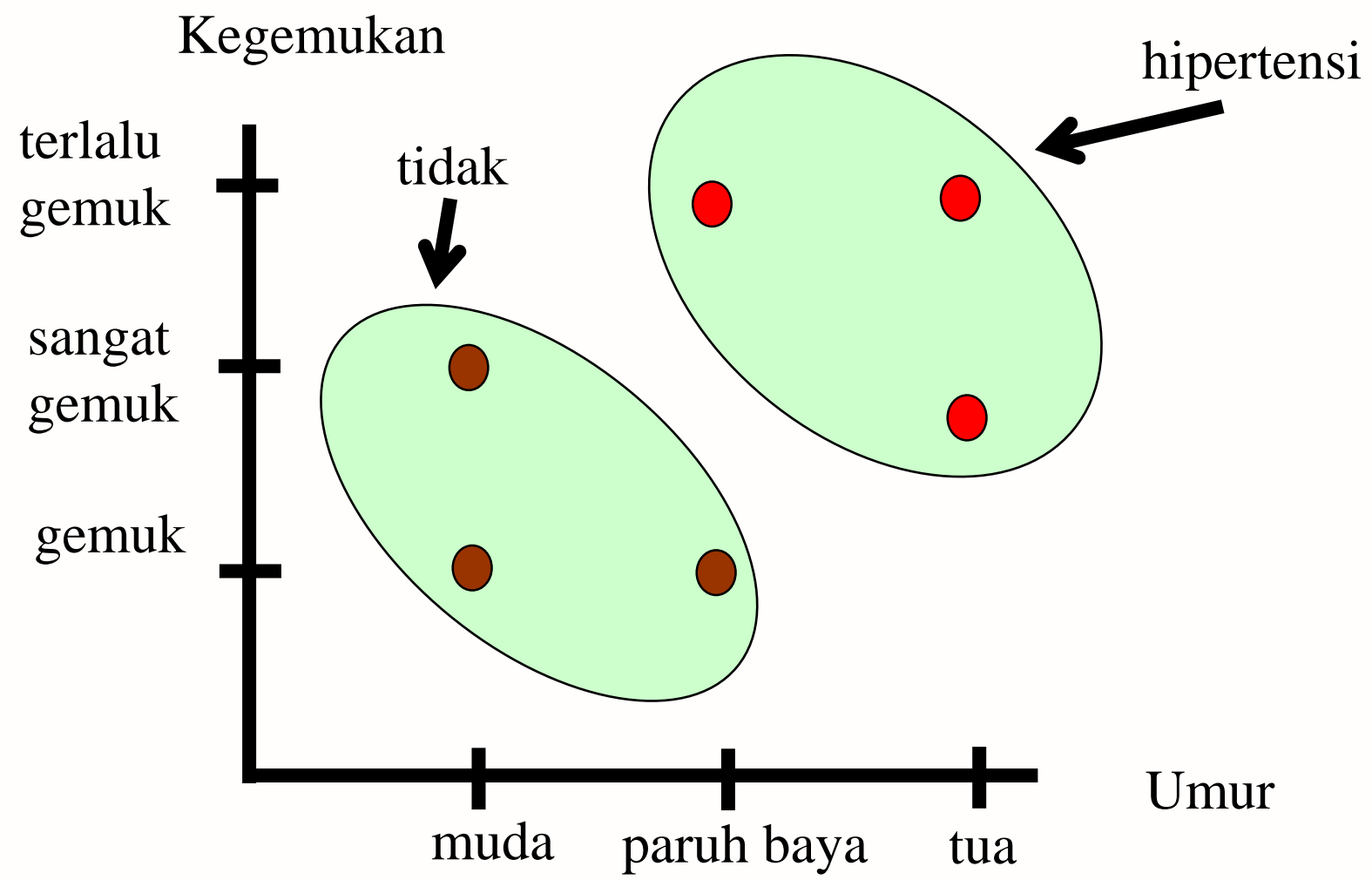


Algoritma k-NN

- Tentukan k
- Hitung jarak antara data baru ke setiap labeled data
- Tentukan k labeled data yang mempunyai jarak yang paling minimal
- Klasifikasikan data baru ke dalam labeled data yang mayoritas

Penyelesaian dengan k-NN (misalnya k=3)





Keuntungan

- Analytically tractable
- Implementasi sangat sederhana
- Tingkat error $>$ bayesian, $<$ 2xbayesian
- Memungkinkan parallel implementation

Kelemahan

- Butuh memori besar
- Komputasi besar

Latihan Soal

1. Implementasikan algoritma 1-NN dalam program!
2. Implementasikan algoritma k-NN dalam program!

Referensi

- Modul Ajar Machine Learning, Entin Martiana, Ali Ridho Barakbah, Nur Rosyid Mubtadaí, Politeknik Elektronika Negeri Surabaya, 2013.
- Machine Learning, Tom Mitchell, McGraw-Hill. 2008.



bridge to the future

<http://www.eepis-its.edu>