

Introduction to Machine Learning

Bahan Kuliah SD3104 Machine Learning

Sevi Nurafni

**Fakultas Sains dan Teknologi
Universitas Koperasi Indonesia 2024**

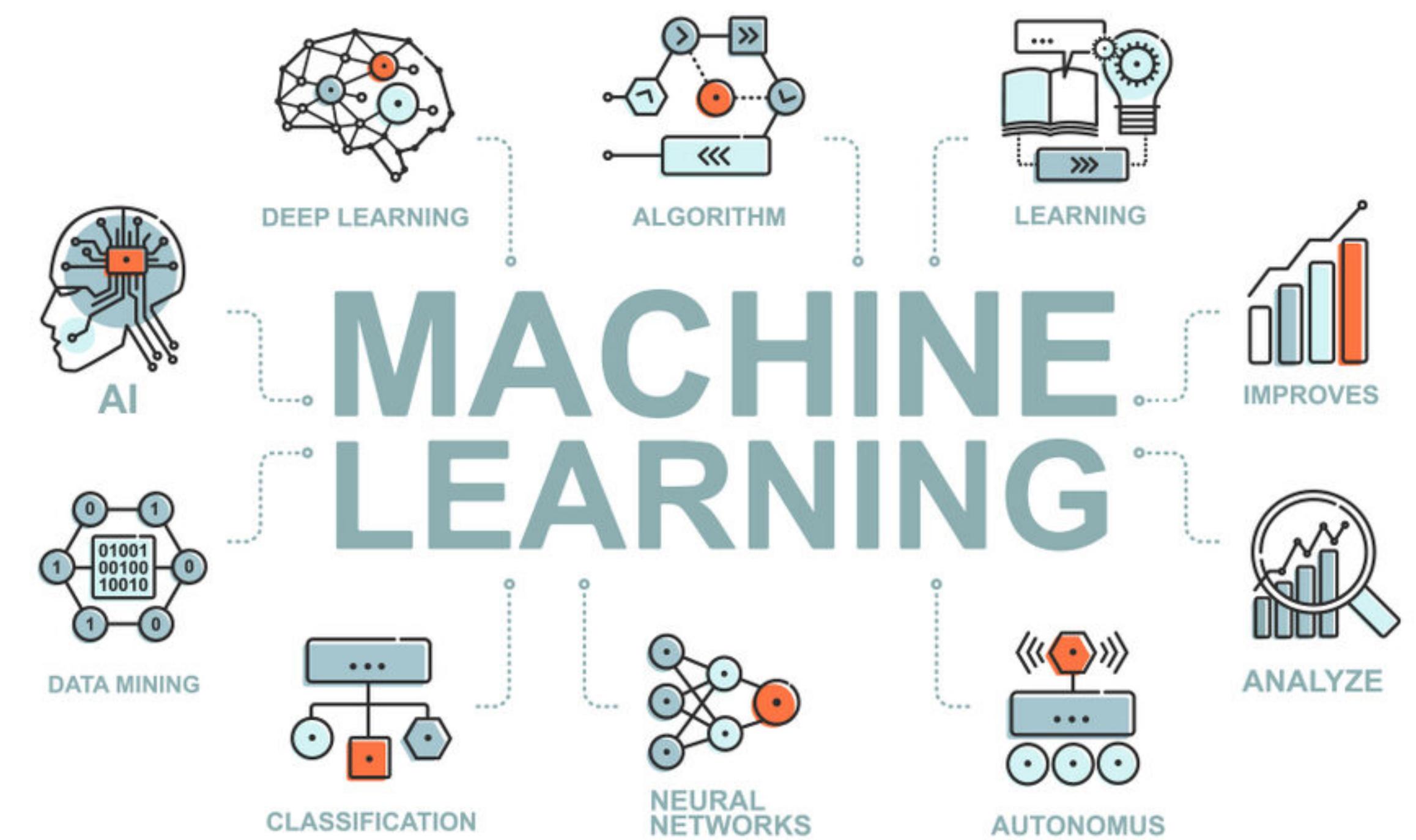
Apa itu Machine Learning?

Definition by Tom Mitchell (1998)

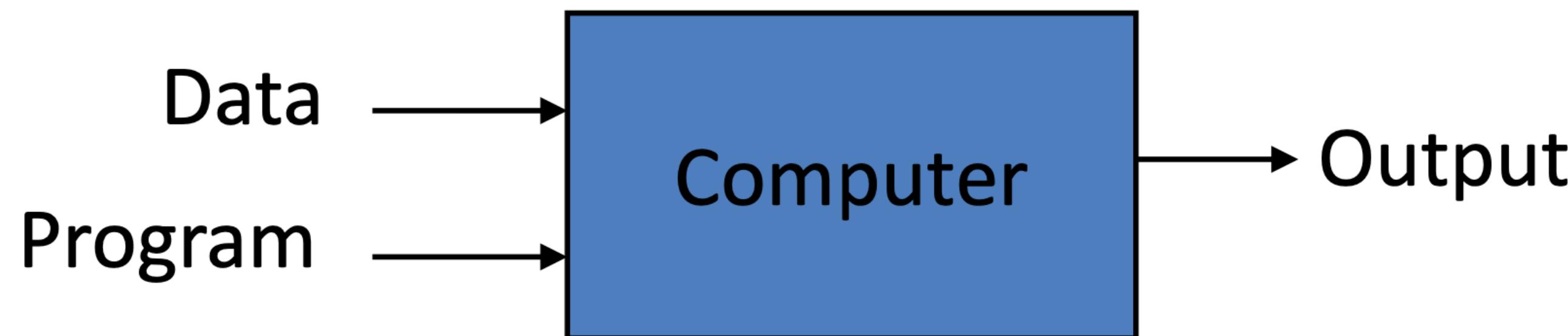
Machine Learning is the study of algorithms that

- Improve their performance P
- at some task T
- with experience E.

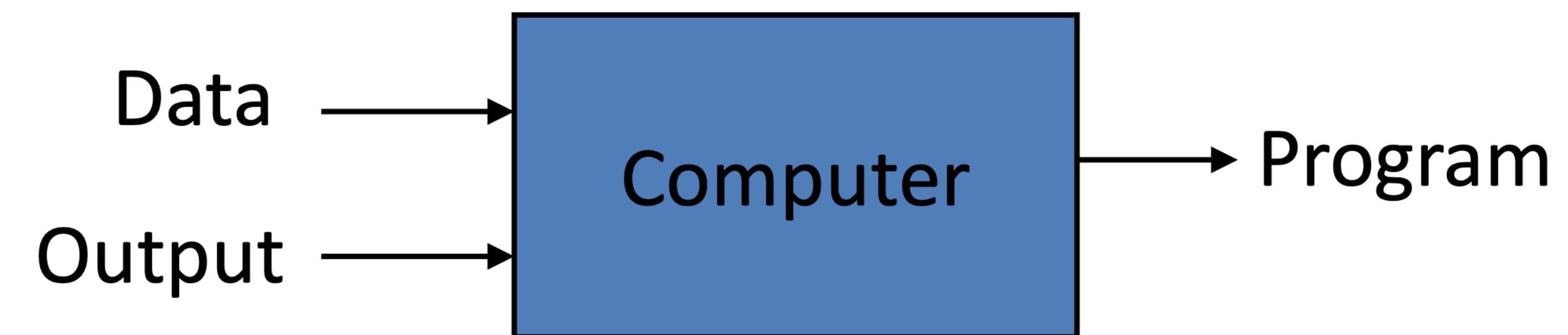
A well-defined learning task is given by $\langle P, T, E \rangle$.



Traditional Programming



Machine Learning



Kapan kita Menggunakan ML?

Pembelajaran Mesin (ML) digunakan ketika:

- Keahlian manusia tidak ada (misalnya navigasi di Mars)
- Manusia tidak dapat menjelaskan keahlian mereka (misalnya pengenalan suara)
- Model harus disesuaikan dengan kebutuhan (misalnya pengobatan yang dipersonalisasi)
- Model didasarkan pada jumlah data yang sangat besar (misalnya genomika)

Learning tidak selalu berguna:

- Tidak perlu "belajar" untuk menghitung gaji

Contoh klasik dari tugas yang memerlukan ML?

Sangat sulit untuk menjelaskan apa yang membuat sebuah angka terlihat seperti angka "2"

0 0 0 1 1 (1 1 1 2
2 2 2 2 2 2 2 3 3 3
3 4 4 4 4 4 5 5 5
4 4 2 2 7 7 7 7 8 8 8
8 8 8 8 9 9 9 9 9

Contoh Tugas yang paling baik diselesaikan dengan menggunakan ML

- Mengenali pola:
 - Identitas wajah atau ekspresi wajah
 - Tulisan tangan atau kata-kata yang diucapkan
 - Gambar medis
- Menghasilkan pola:
 - Menghasilkan gambar atau urutan gerakan
- Mengenali anomali:
 - Transaksi kartu kredit yang tidak biasa
 - Pola bacaan sensor yang tidak biasa di pembangkit listrik tenaga nuklir
- Prediksi:
 - Harga saham atau nilai tukar mata uang di masa depan

Contoh Penerapan

- Biologi komputasional
- Keuangan
- E-commerce
- Penjelajahan luar angkasa
- Robotika
- Ekstraksi informasi
- Jejaring sosial
- Debugging perangkat lunak
- [Bidang favorit Anda]

Samuel's Checkers-Player

“Machine Learning: Bidang studi yang memberi komputer kemampuan untuk belajar tanpa diprogram secara eksplisit.”

— Arthur Samuel (1959)



Improve on task T, with respect to performance metric P, based on experience E



T: Mengenali kata-kata yang ditulis tangan

P: Persentase kata yang diklasifikasikan dengan benar

E: Basis data gambar kata-kata tulisan tangan yang diberi label oleh manusia

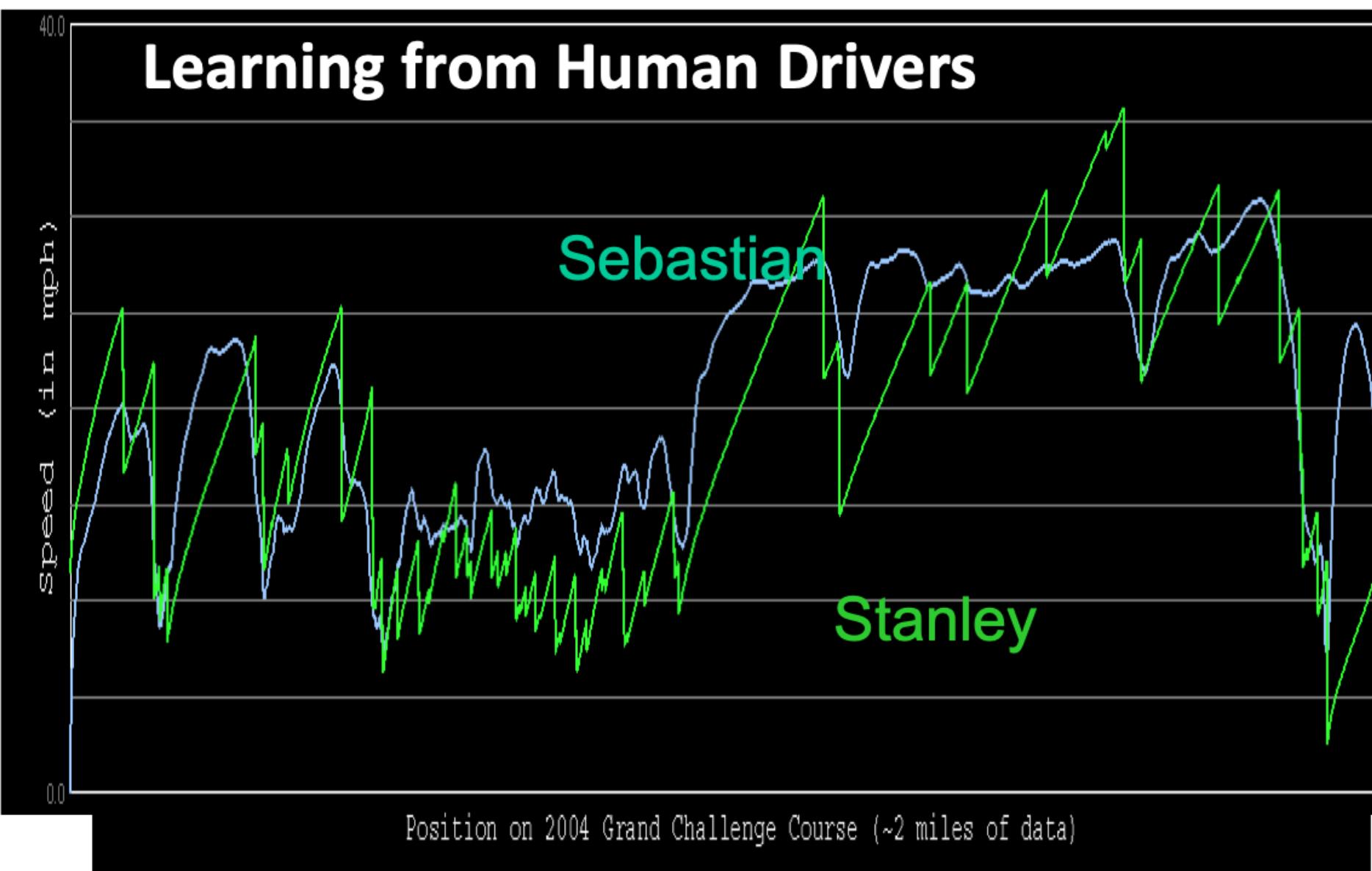
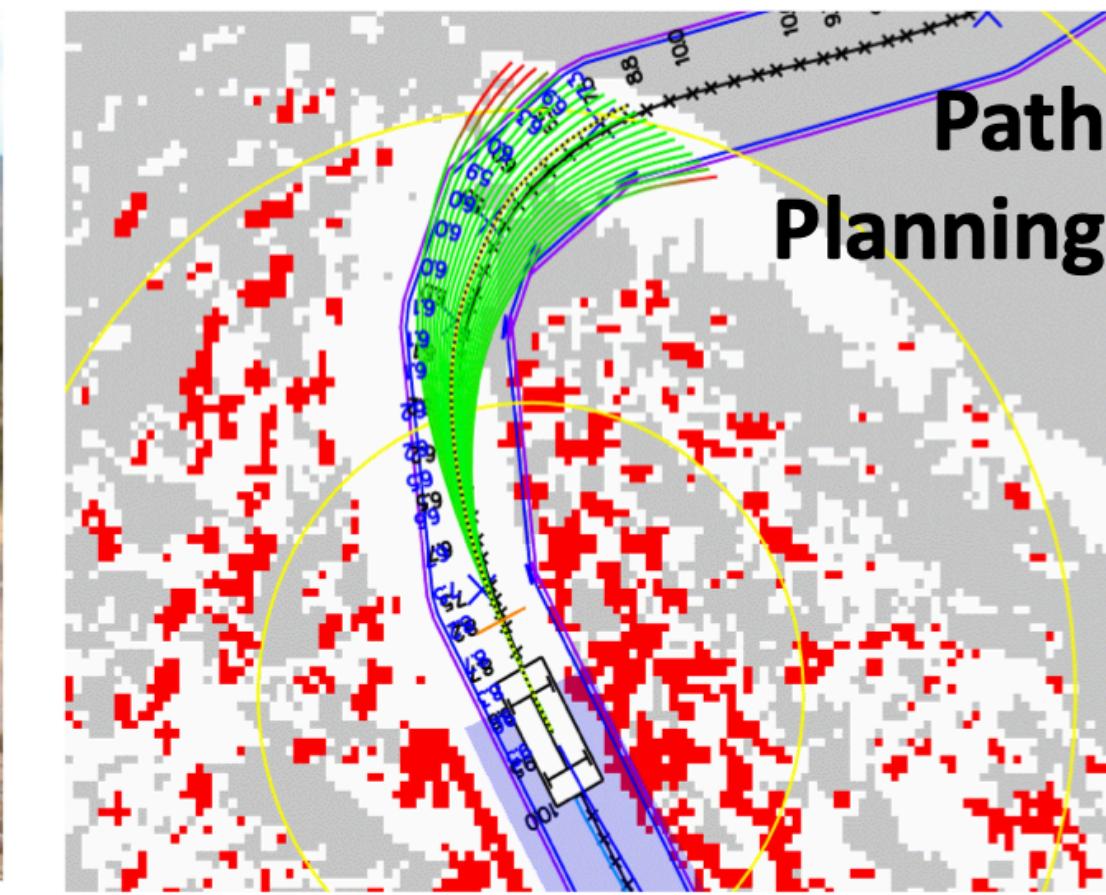
T: Mengategorikan email sebagai spam atau sah

P: Persentase email yang diklasifikasikan dengan benar

E: Basis data email, sebagian dengan label yang diberikan oleh manusia

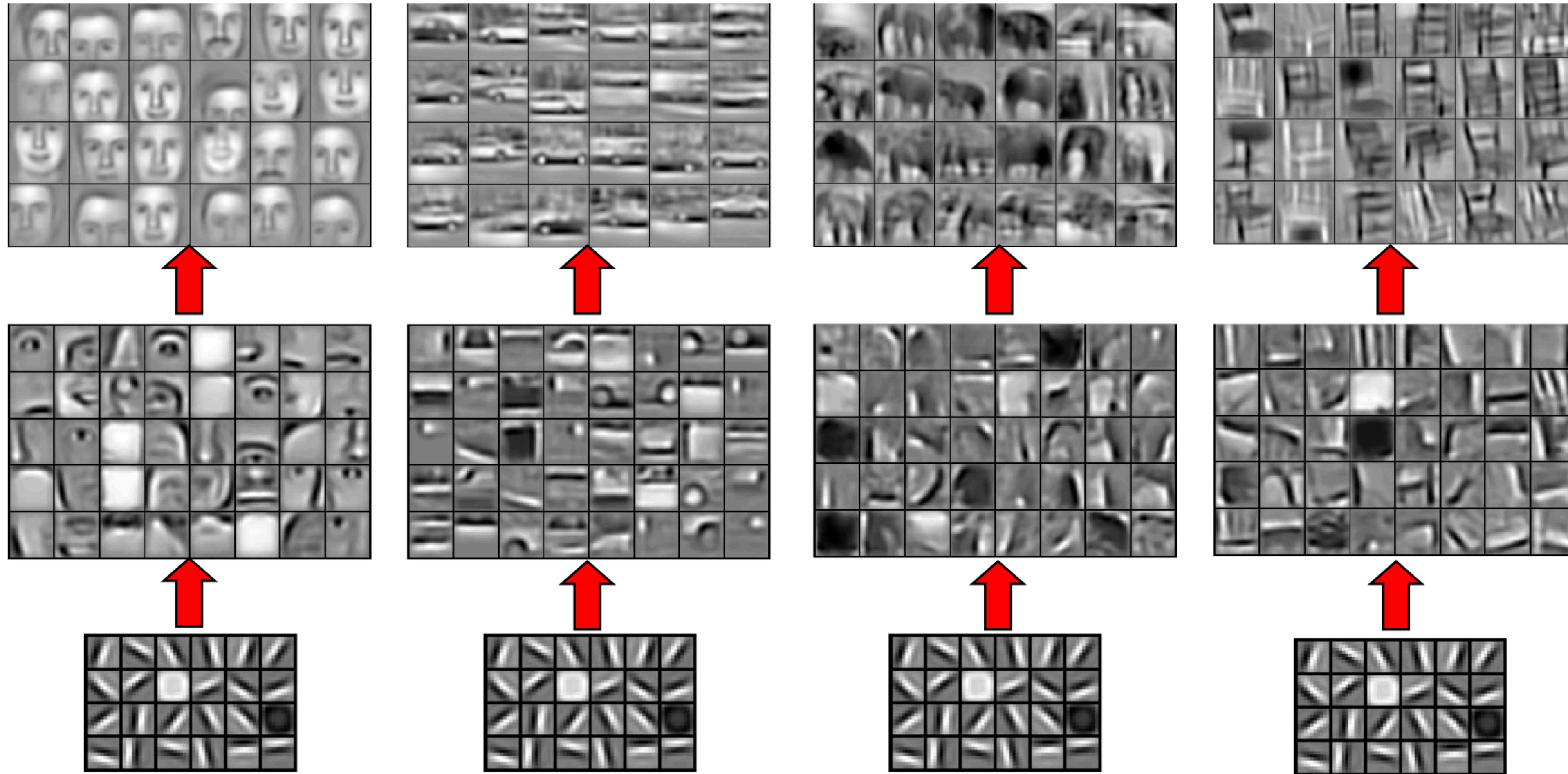
Aplikasi Terkini dari Pembelajaran Mesin

Autonomous Car Technology



Images and movies taken from Sebastian Thrun's multimedia website.

Learning dari Bagian-bagian Objek



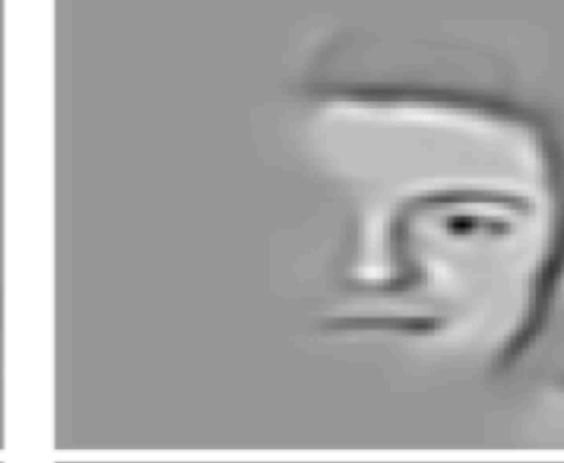
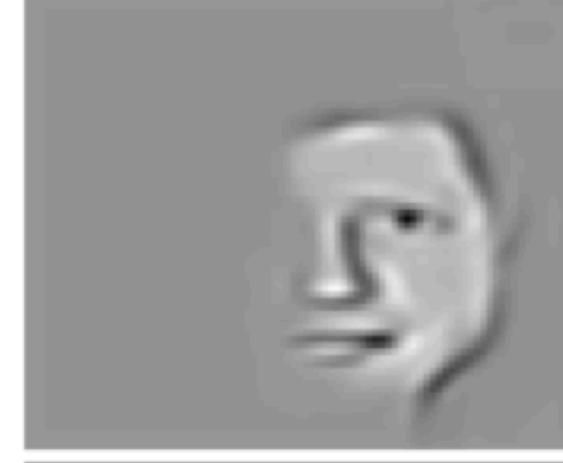
Inferensi dari Model Deep Learning

Menghasilkan sampel posterior dari wajah dengan eksperimen "mengisi" (cf. Lee dan Mumford, 2003). Menyajikan inferensi dari bawah ke atas dan dari atas ke bawah.

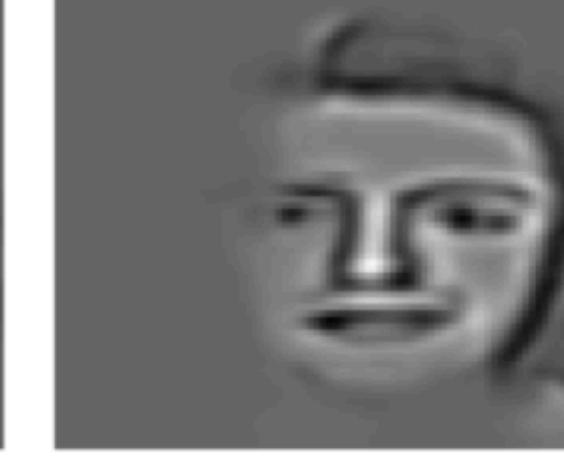
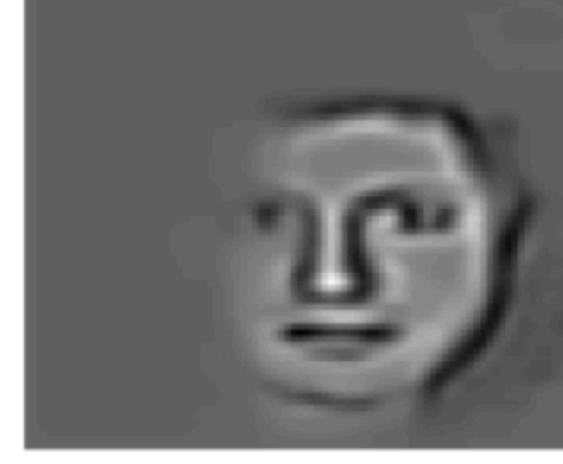
Input images



Samples from
feedforward
Inference
(control)



Samples from
Full posterior
inference



Jenis-Jenis dari Learning

Supervised Learning (induktif)

Diberikan: data pelatihan + keluaran yang diinginkan (label)

Unsupervised Learning

Diberikan: data pelatihan (tanpa keluaran yang diinginkan)

Semi-supervised Learning

Diberikan: data pelatihan + beberapa keluaran yang diinginkan

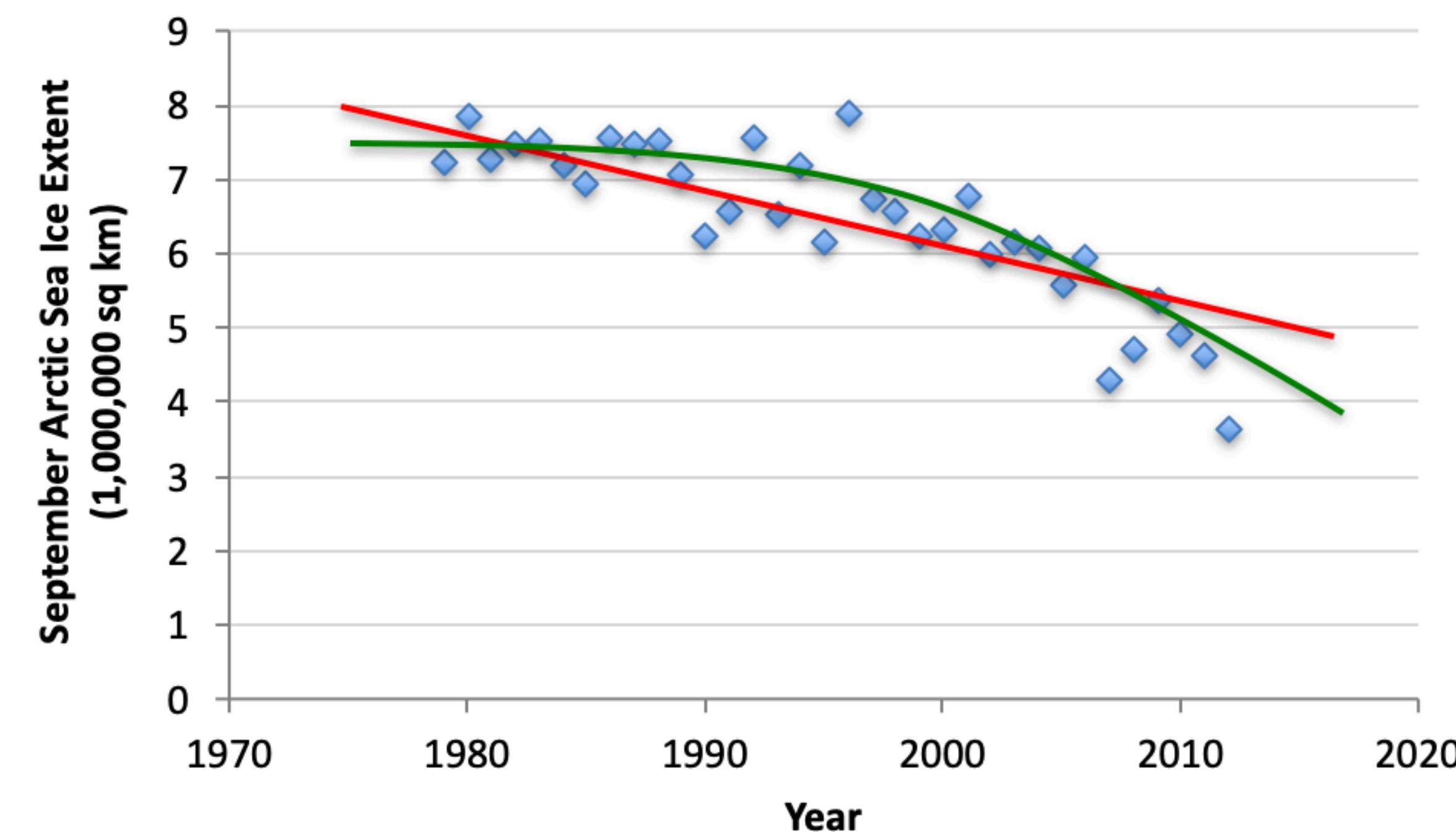
Reinforcement Learning

Sistem belajar dari tindakan-tindakannya berdasarkan umpan balik yang diterima dari setiap tindakan

Supervised Learning: Regression

Diberikan $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

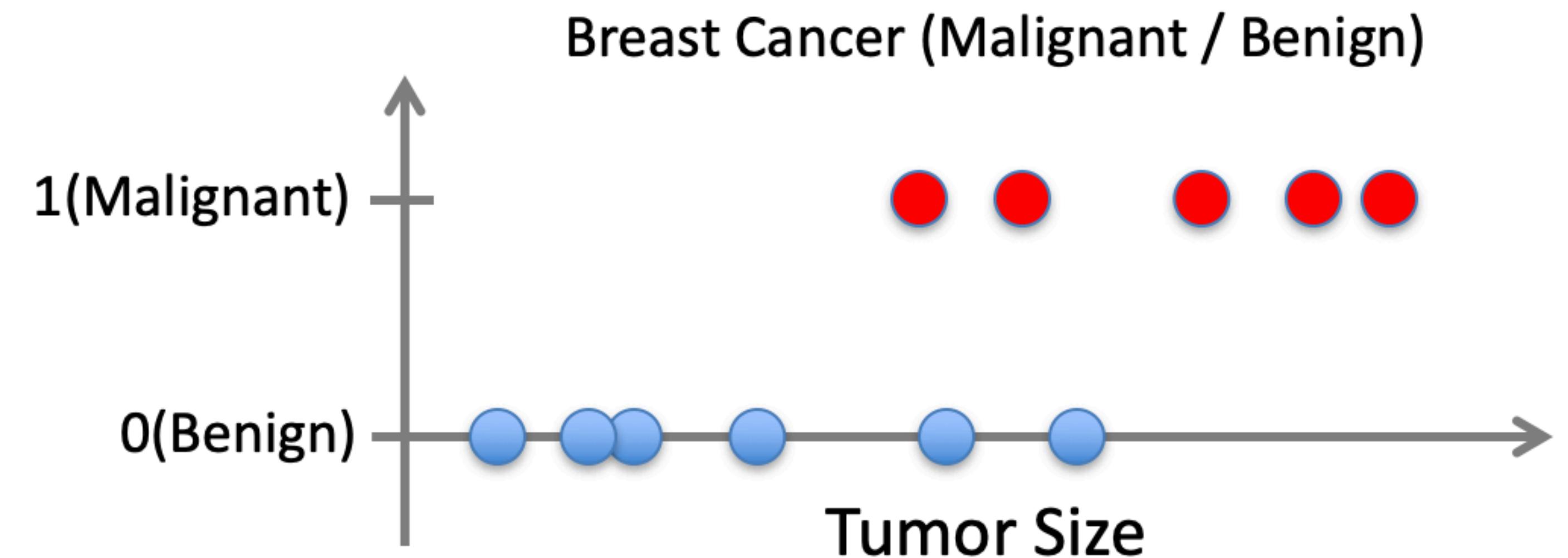
Mempelajari fungsi $f(x)$ untuk memprediksi y yang diberikan x
 y bernilai riil == regresi



Supervised Learning: Classification

Diberikan $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

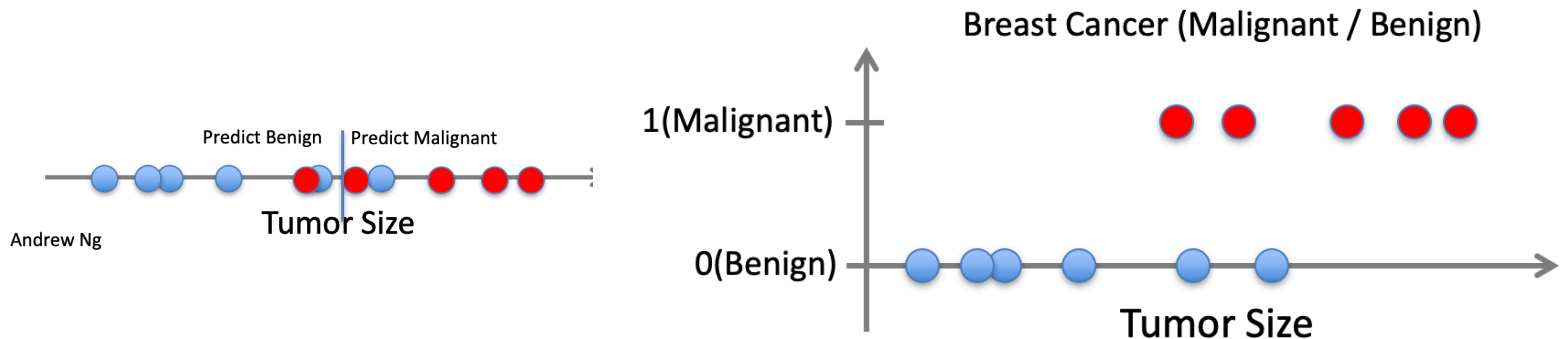
Mempelajari fungsi $f(x)$ untuk memprediksi y yang diberikan x
 y adalah kategorikal == klasifikasi



Supervised Learning: Classification

Diberikan $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

Mempelajari fungsi $f(x)$ untuk memprediksi y yang diberikan x
 y adalah kategorikal == klasifikasi



Supervised Learning

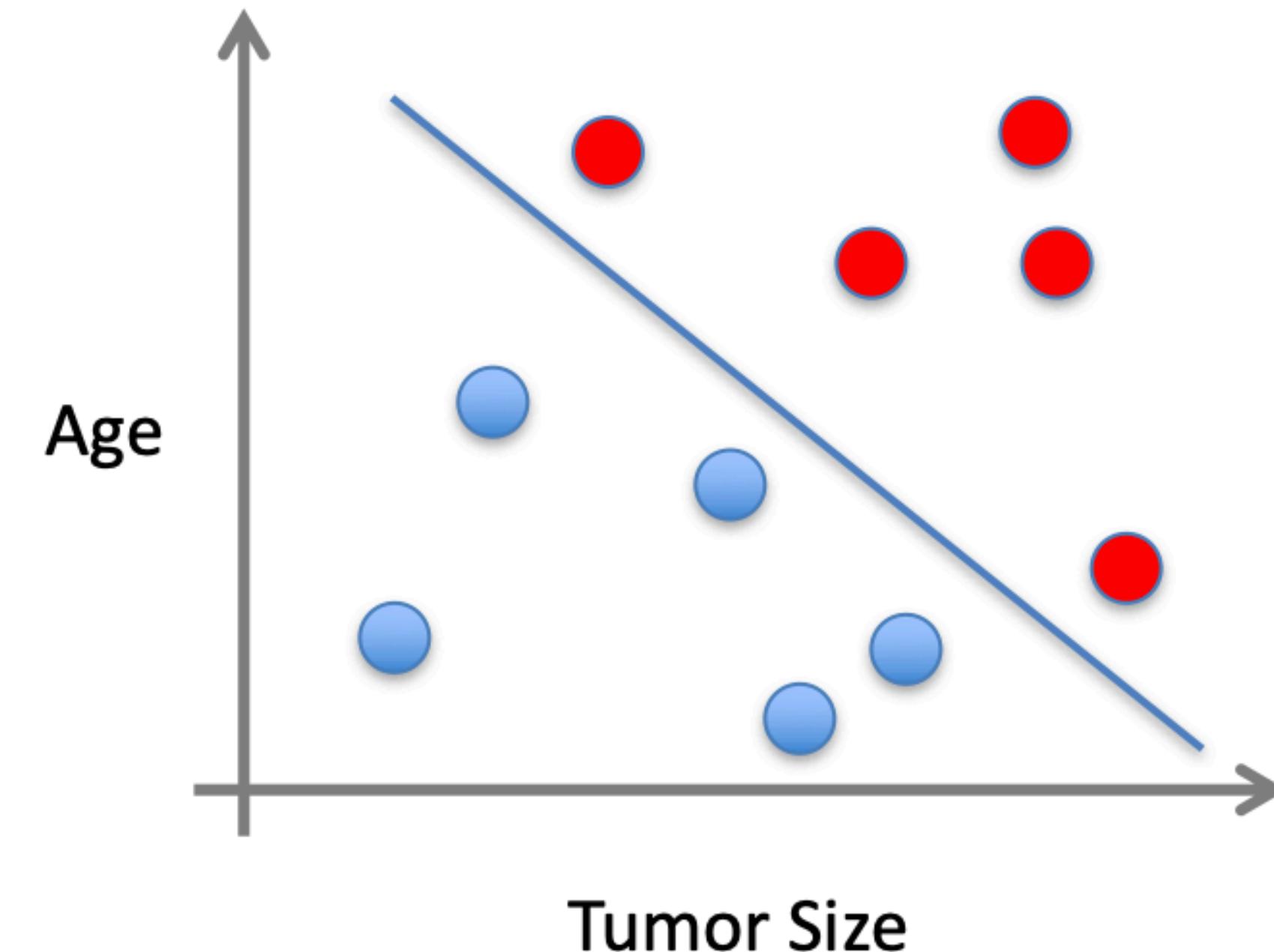
x dapat berupa multi-dimensi

- Setiap dimensi berhubungan dengan sebuah atribut

Ketebalan Gumpalan

Keseragaman ukuran Sel

Keseragaman Bentuk Sel

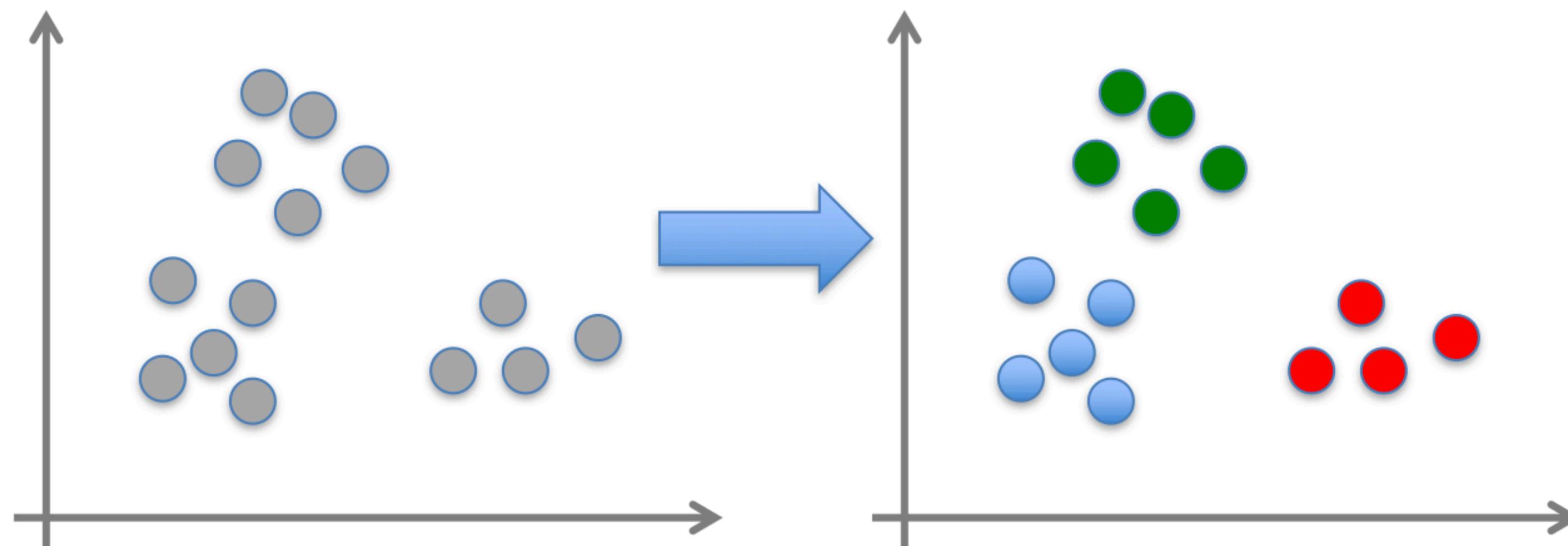


Unsupervised Learning

Diberikan x_1, x_2, \dots, x_n (tanpa label)

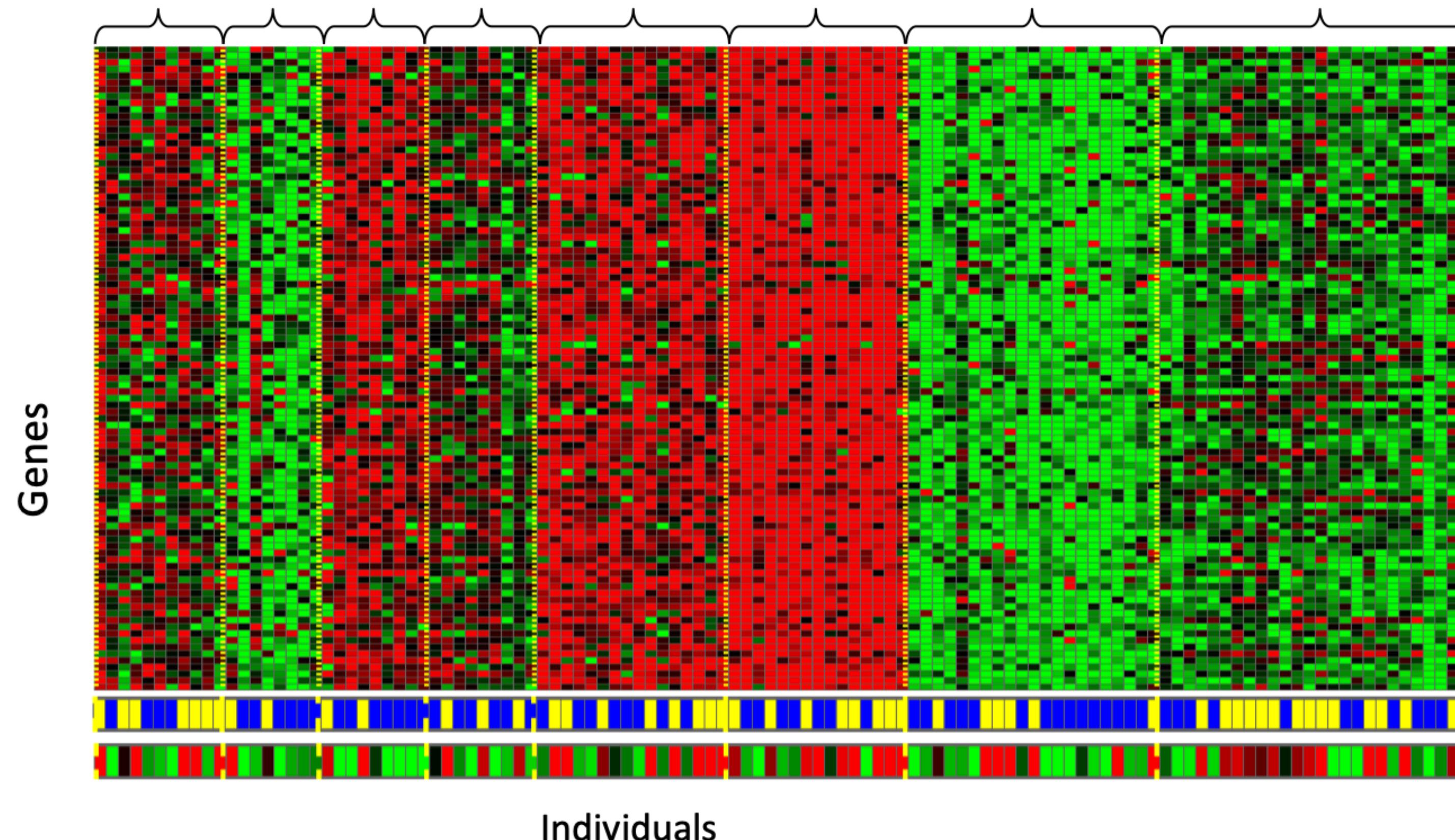
Outputnya struktur tersembunyi di balik x

Clustering



Unsupervised Learning

Aplikasi genomik: mengelompokkan individu berdasarkan kesamaan genetik



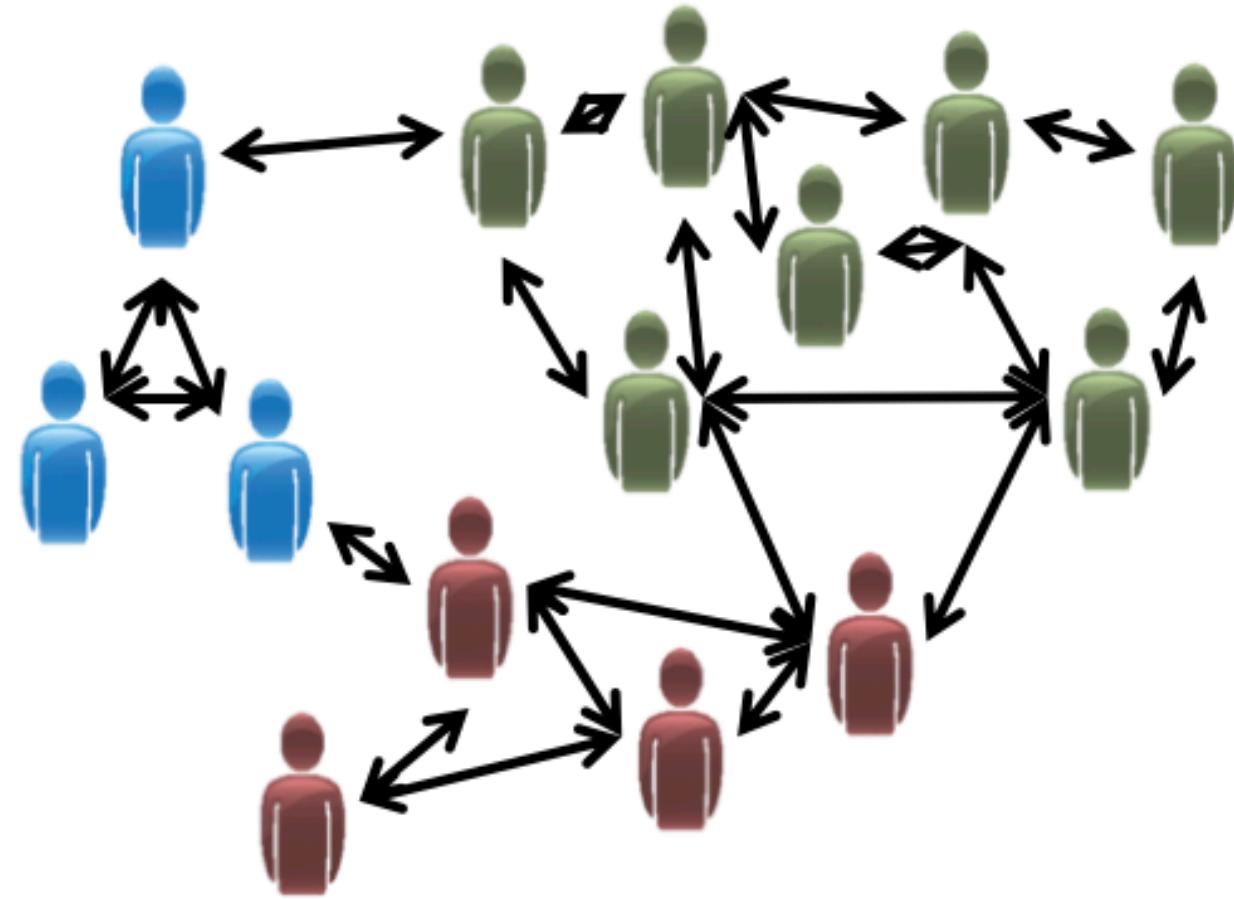
Unsupervised Learning



Organize computing clusters



Market segmentation



Social network analysis



Astronomical data analysis

Reinforcement Learning

Diberikan serangkaian kondisi (state) dan tindakan, dengan penghargaan yang mungkin muncul belakangan, hasilkan sebuah strategi (policy).

– **Strategi (policy)** adalah aturan yang menghubungkan kondisi tertentu dengan tindakan yang harus dilakukan. Aturan ini memberi tahu apa yang harus dilakukan dalam kondisi tertentu.

Contoh:

- Bermain game
- Robot yang harus keluar dari labirin
- Menyeimbangkan tongkat di tangan

Reinforcement Learning

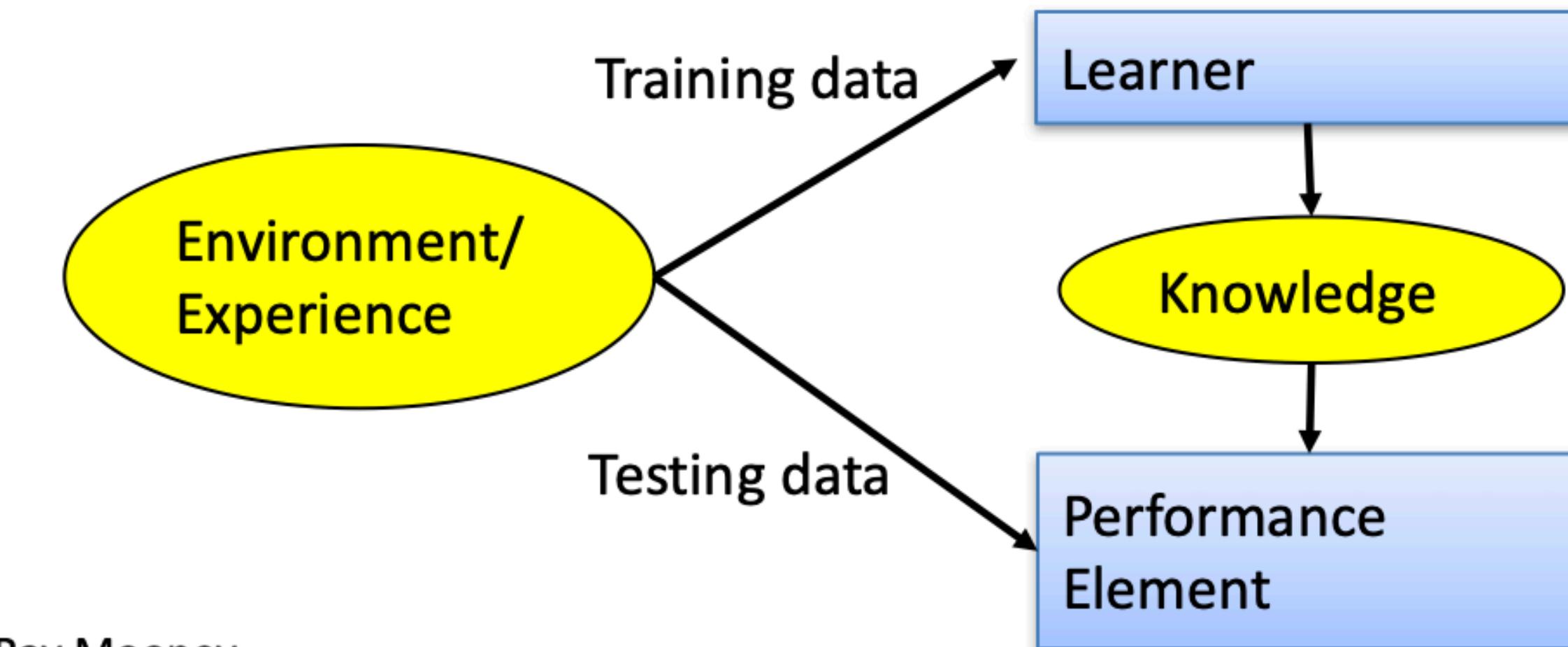


<https://www.youtube.com/watch?v=4cgWya-wjgY>

Merumuskan Masalah dari Learning

Merancang Sistem Learning

- Pilih experience training yang tepat
- Pilih secara spesifik apa yang ingin dipelajari – yaitu fungsi targetnya
- Pilih bagaimana cara merepresentasikan fungsi target
- Pilih algoritma pembelajaran untuk menyimpulkan fungsi target dari pengalaman tersebut



by Ray Mooney

Training vs. Test Distribution

Secara umum, kita mengasumsikan bahwa contoh-contoh pelatihan dan pengujian diambil secara independen dari distribusi data yang sama.

- Ini disebut "i.i.d" yang merupakan singkatan dari "independent and identically distributed" (independen dan terdistribusi identik).
 - Jika contoh-contoh tersebut tidak independen, maka kita membutuhkan klasifikasi kolektif.
 - Artinya, contoh-contoh saling terkait, sehingga kita harus mempertimbangkan hubungan antarcontoh saat melakukan klasifikasi. Contohnya, dalam jaringan sosial, aktivitas pengguna mungkin bergantung pada aktivitas pengguna lain.
 - Jika distribusi data pengujian berbeda dari distribusi data pelatihan, kita membutuhkan transfer learning.
 - Ini berarti kita harus menerapkan model yang dilatih di satu lingkungan (data pelatihan) ke lingkungan baru yang berbeda (data pengujian), dengan menyesuaikan model agar dapat berfungsi dengan baik pada distribusi data yang berbeda.

ML in a Nutshell

Ada puluhan ribu algoritma pembelajaran mesin (machine learning)

– Ratusan algoritma baru muncul setiap tahun.

Setiap algoritma ML memiliki tiga komponen:

- Representasi
- Optimasi
- Evaluasi

Berbagai Representasi Fungsi

- Numerical functions
 - Linear regression
 - Neural networks
 - Support vector machines
- Symbolic functions – Decision trees
 - Rules in propositional logic
 - Rules in first-order predicate logic
- Instance-based functions – Nearest-neighbor
 - Case-based
- Probabilistic Graphical Models – Naïve Bayes
 - Bayesian networks
 - Hidden-Markov Models (HMMs)
 - Probabilistic Context Free Grammars (PCFGs) – Markov networks

Berbagai Optimasi Algoritma

- Gradient descent
 - Perceptron
 - Backpropagation
- Dynamic Programming
 - HMMLearning
 - PCFGLearning
- Divide and Conquer
 - Decisiontreeinduction
 - Rulelearning
- Evolutionary Computation
 - GeneticAlgorithms(GAs)
 - GeneticProgramming(GP)
 - Neuro-evolution

Evaluasi

Accuracy

- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost/Utility
- Margin
- Entropy
- K-L divergence
- etc.

ML in Practice



Pahami domain, pengetahuan sebelumnya, dan tujuan
Data integration, selection, cleaning, pre-processing, etc.
Learn models
Menginterpretasikan hasil
Konsolidasi dan penerapan pengetahuan yang ditemukan

Pelajaran yang Dipetik tentang Learning

- Pembelajaran dapat dilihat sebagai penggunaan pengalaman langsung atau tidak langsung untuk mendekati fungsi target yang dipilih.
- Aproksimasi fungsi dapat dilihat sebagai pencarian melalui ruang hipotesis (representasi fungsi) untuk menemukan yang paling cocok dengan sekumpulan data pelatihan.
- Metode pembelajaran yang berbeda memiliki ruang hipotesis (bahasa representasi) yang berbeda dan/atau menggunakan teknik optimasi yang berbeda.

Apa yang Akan Kita Bahas?

Supervised learning

- Decision tree induction
 - – Linear regression
 - – Logistic regression
 - – Support vector machines & kernel methods
 - – Model ensembles
 - – Bayesian learning
 - – Neural networks & deep learning
 - – Learning theory
-

Unsupervised learning – Clustering

- Dimensionality reduction
- **Reinforcement learning**
 - Temporal difference learning
 - Q learning
 - **Evaluation**
 - **Applications**
 -

**SELAMAT
BELAJAR**