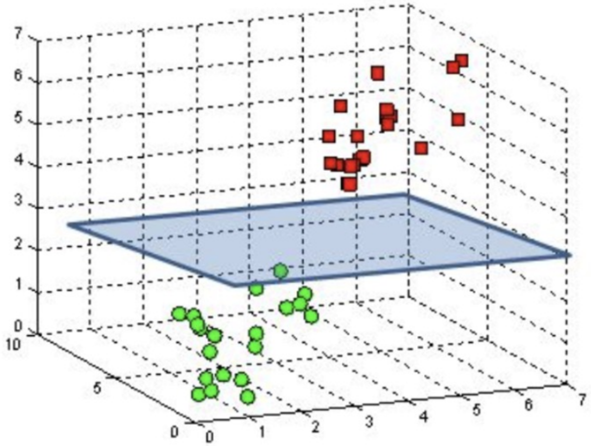
DEPARTEMEN
TEKNIK INFORMATIKA

Support Vector Machine



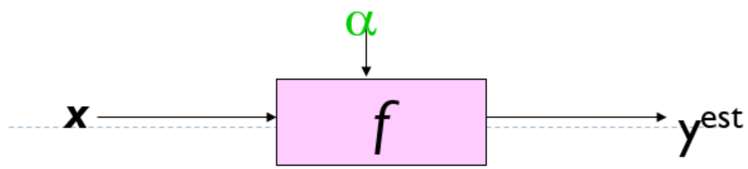
Kecerdasan Komputasional

Dini Adni Navastara

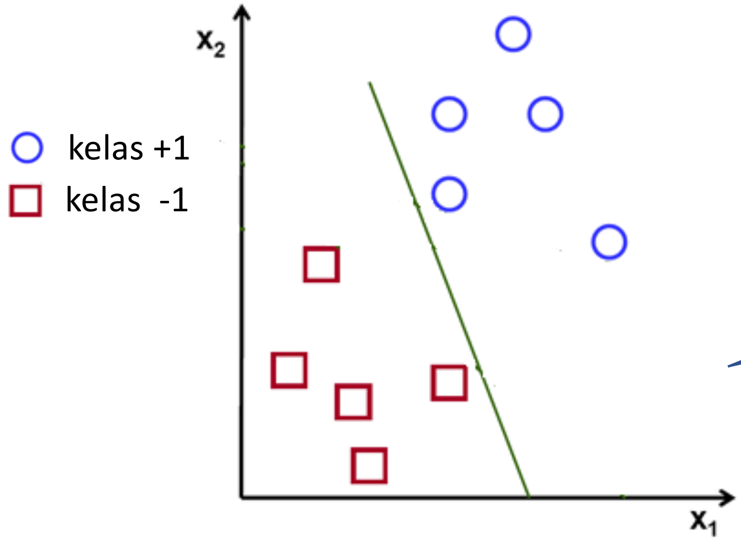
Capaian Pembelajaran

Mahasiswa mampu menjelaskan konsep klasifier dengan fungsi diskriminan linear maupun non-linear

Linear Classifiers

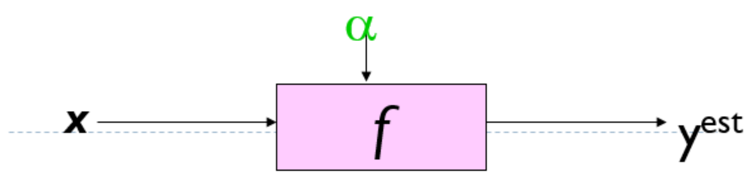


$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} - b)$

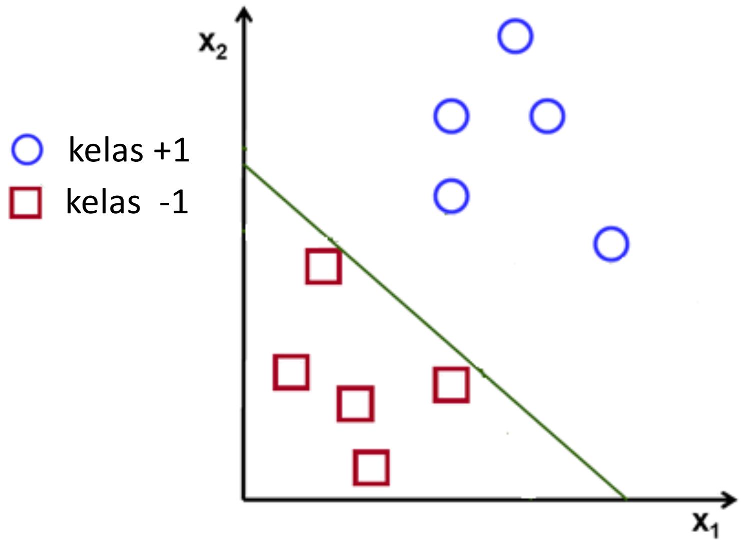


Bagaimana mengklasifikasikan data ini?

Linear Classifiers

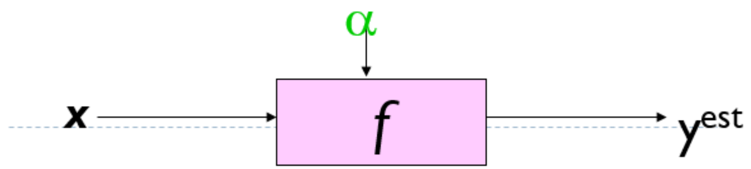


$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w} \cdot \mathbf{x} - b)$

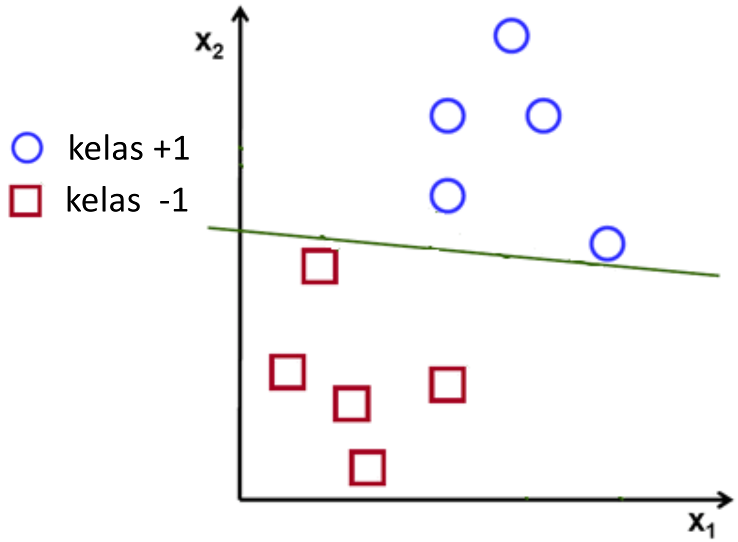


Bagaimana mengklasifikasikan data ini?

Linear Classifiers

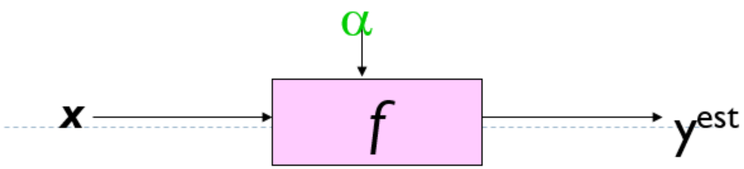


$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} - b)$

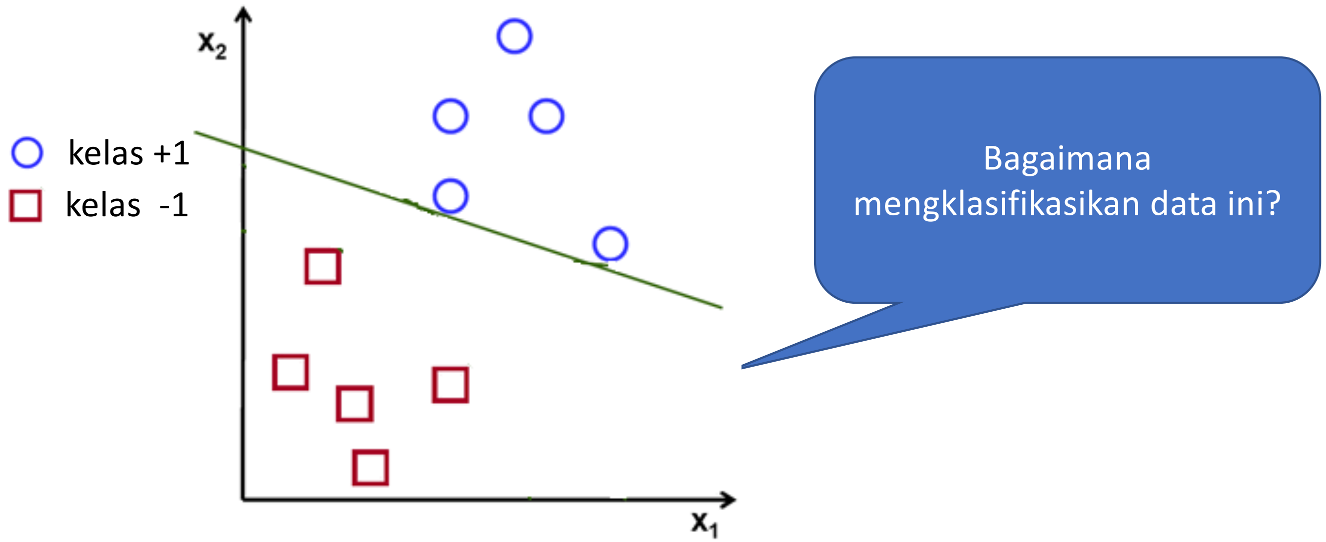


Bagaimana mengklasifikasikan data ini?

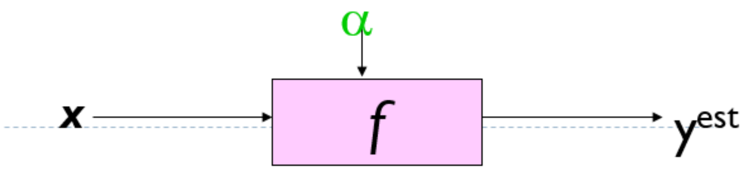
Linear Classifiers



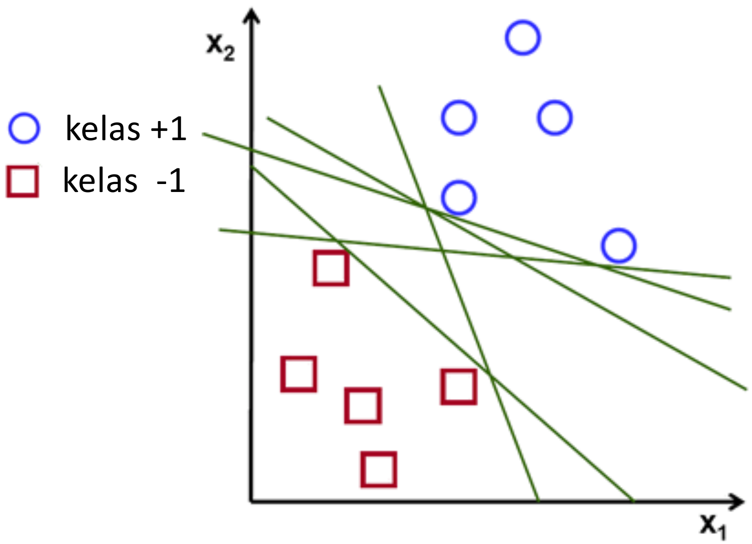
$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} - b)$



Linear Classifiers



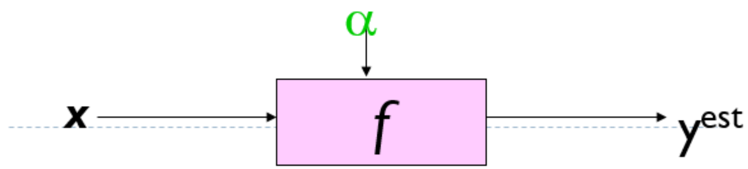
$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} - b)$



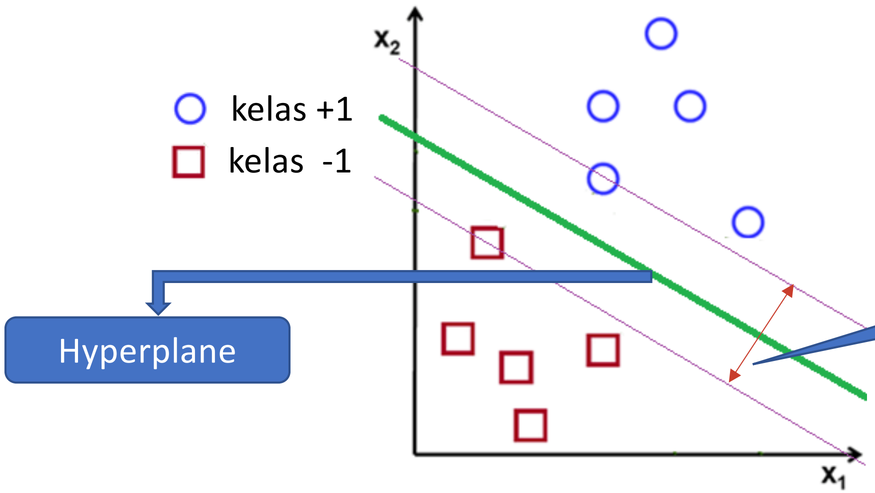
Yang mana yang terbaik?
(Hyperplane terbaik)?

Classifiers Margin

Ide dasar SVM: **memaksimalkan distance (jarak) antara *hyperplane* dan titik sampel terdekat**

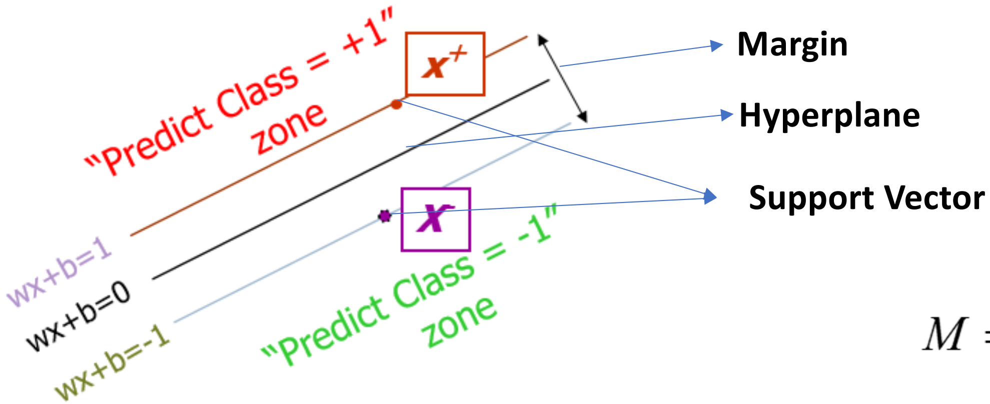


$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} - b)$



Definisikan **margin/batas** linear classifier sebagai lebarnya dimana batas tersebut dapat dimaksimalkan mencapai titik data terdekat

Linear SVM secara Matematis



$$M = \frac{(x^+ - x^-) \cdot w}{|w|} = \frac{2}{|w|}$$

- $\boldsymbol{w} \,.\, \boldsymbol{x^+} + b = +1$
- $\boldsymbol{w} \,.\, \boldsymbol{x^-} + b = -1$
- $\boldsymbol{w} \,.\, (\boldsymbol{x^+} - \boldsymbol{x^-}) = 2$

Dimana,
M: Margin
w: bobot
x^+: support vector kelas +1
x^-: support vector kelas -1

Problem Optimasi

- $\{x_1, \ldots, x_n\}$ adalah data set dan $y_i \in \{1, -1\}$ adalah kelas label dari $\boldsymbol{x_i}$
- Batas keputusan harus dapat mengklasifikasi semua titik dengan benar

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, \qquad \forall i$$

- Problem optimasi terbatas

meminimalkan $\frac{1}{2}||\mathbf{w}||^2$

dimana $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \qquad \forall i$

Problem Optimasi

- Kita dapat mengubah problem menjadi bentuk dual

$$\text{max.} \quad W(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

dimana $\alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i y_i = 0$

- Problem *quadratic programming* (QP)
  - Global maximum pada α_i dapat selalu ditemukan
- $\boldsymbol{w}$ dapat diperbaiki menjadi $\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$

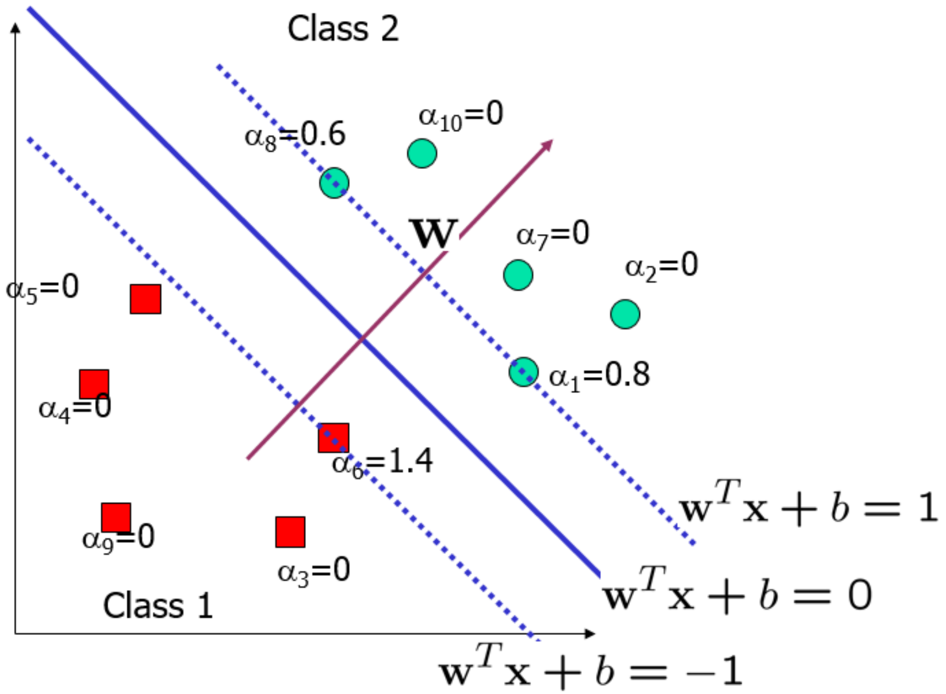
Karakteristik Solusi

- Kebanyakan nilai α_i adalah nol
- $\boldsymbol{x_i}$ dengan nilai α_i positif disebut sebagai *support vectors* (SV)
  - Batas keputusan ditentukan oleh SV
  - Diketahui t_j (j=1, ..., s) adalah indeks dari SV

$$\mathbf{w} = \sum_{j=1}^{s} \alpha_{t_j} y_{t_j} \mathbf{x}_{t_j}$$

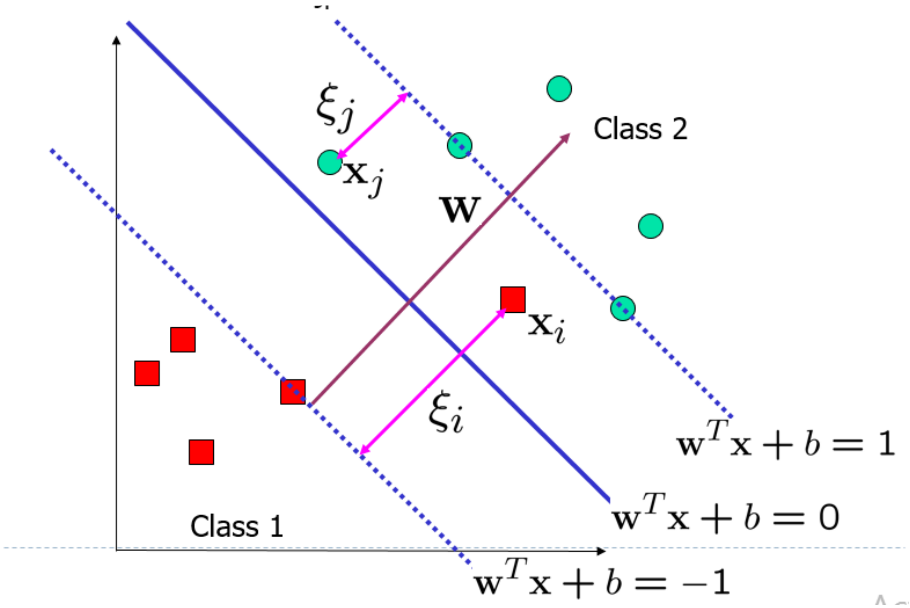
- Untuk testing dengan data baru ($\boldsymbol{z}$)
  - Hitung $\mathbf{w}^T\mathbf{z} + b = \sum_{j=1}^{s} \alpha_{t_j} y_{t_j} (\mathbf{x}_{t_j}^T \mathbf{z}) + b$
  - Dan klasifikasikan $\boldsymbol{z}$ sebagai kelas 1 jika jumlahnya positif, selain itu diklasifikasikan sebagai kelas 2

Interpretasi Geometri



Tidak dapat Dipisahkan secara Linier?

- Memperbolehkan adanya *error* (ξ) pada klasifikasi



Soft Margin Hyperplane

- Definisikan ξ_i=0 jika tidak ada error untuk $\boldsymbol{x_i}$
  - ξ_i adalah variabel untuk mendefinisikan nilai error

$$\begin{cases} \mathbf{w}^T\mathbf{x}_i + b \geq 1 - \xi_i & y_i = 1 \\ \mathbf{w}^T\mathbf{x}_i + b \leq -1 + \xi_i & y_i = -1 \\ \xi_i \geq 0 & \forall i \end{cases}$$

- Minimalkan $\frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i$
  - *C* : tradeoff parameter antara error dan margin
- Problem optimasi menjadi

meminimalkan $\frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i$

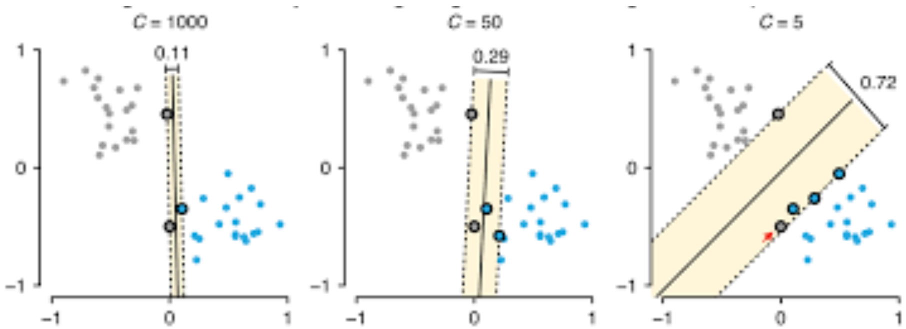
dimana $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$

Soft Margin Hyperplane

- Menemukan nilai yang tepat untuk *C* menjadi salah satu masalah dalam SVM

Nilai ***C*** berperan dalam mengontrol *overfitting*.

- ***C*** besar → lebih sedikit sampel training yang berada di posisi yang tidak ideal (artinya lebih sedikit error, sehingga berdampak positif pada kinerja *classifier*) → **C terlalu besar menyebabkan *overfitting***
- ***C*** kecil → lebih banyak sampel training yang tidak berada pada posisi ideal (artinya akan banyak error training sehingga berdampak negatif pada kinerja *classifier*) → **C terlalu kecil menyebabkan *underfitting***



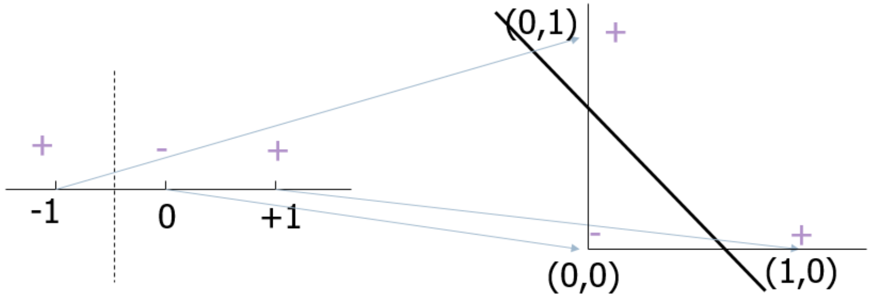
Permasalahan Non-linear

- Ide: transformasi $\boldsymbol{x}_i$ ke ruang berdimensi lebih tinggi untuk memudahkan perhitungan
  - Ruang Input : ruang $\boldsymbol{x}_i$
  - Ruang Fitur: ruang $\phi(\boldsymbol{x}_i)$ setelah transformasi
- Mengapa perlu transformasi?
  - Operasi linear pada ruang fitur ekivalen dengan operasi non-linear pada ruang input
  - Proses klasifikasi lebih mudah dilakukan dengan transformasi.

  Contoh: XOR

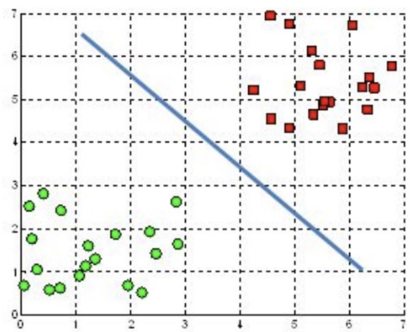
Dimensi Tinggi

- Proyeksikan data ke ruang berdimensi tinggi agar data-data tersebut dapat dipisahkan secara linear dan dapat menggunakan linear SVM – (Using Kernels)

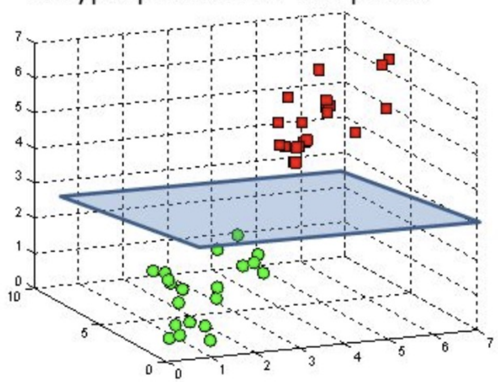


Data dari $\mathbf{R}^1$ ditransformasi ke $\mathbf{R}^2$

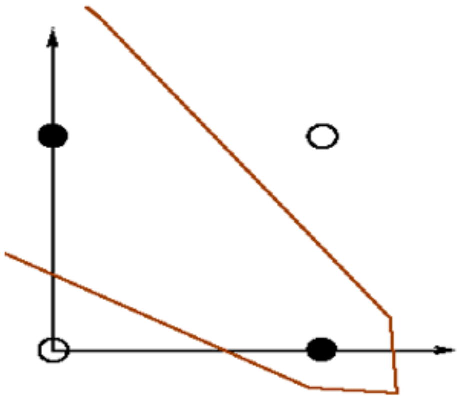
A hyperplane in $\mathbb{R}^2$ is a line



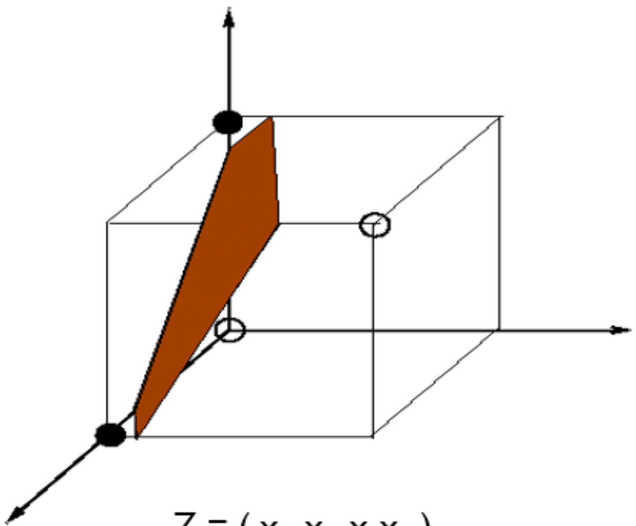
A hyperplane in $\mathbb{R}^3$ is a plane



Problem XOR



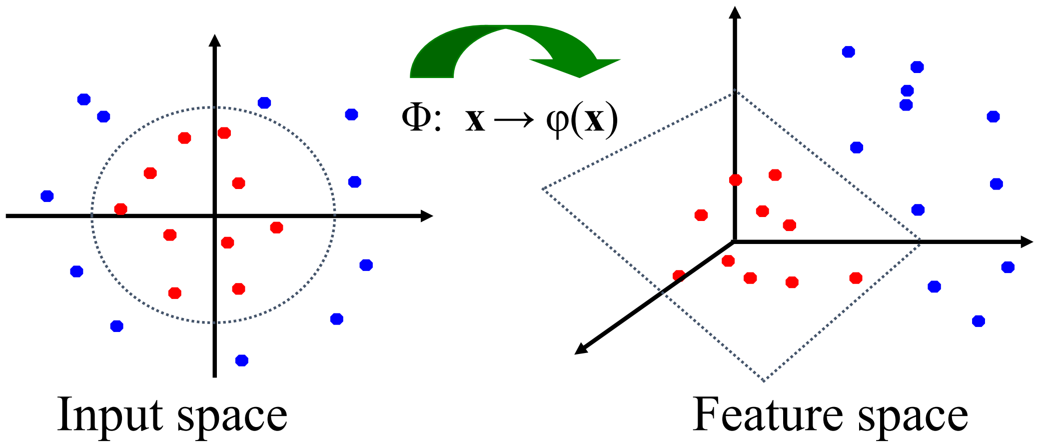
$X = (x_1, x_2)$



$Z = (x_1, x_2, x_1 x_2)$

Permasalahan Non-linear

- Kemungkinan problem transformasi
  - Komputasi yang tinggi dan sulit memperoleh estimasi bagus
- SVM menyelesaikan masalah ini secara bersamaan
  - Kernel tricks untuk komputasi yang efisien
  - Meminimalkan $||\boldsymbol{w}||^2$ dapat menghasilkan classifier yang baik



Kernel Trick

- Hubungan antara fungsi kernel K dan mapping $\phi(.)$

$$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y}) = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{y}) \rangle$$

  - Disebut sebagai **kernel trick**

- Secara intuitif, $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y})$ merepresentasikan kemiripan antara data $\boldsymbol{x}$ dan $\boldsymbol{y}$, dan ini diperoleh dari pengetahuan sebelumnya

Jenis-jenis Fungsi Kernel

No	Nama Kernel	Definisi Fungsi
1	Linier	$K(x,y) = x.y$
2	Polinomial of degree d	$K(x,y) = (x.y)^d$
3	Polinomial of degree up to d	$K(x,y) = (x.y + c)^d$
4	Gaussian RBF	$K(x,y) = \exp\left(\frac{-\lVert x-y \rVert^2}{2\sigma^2}\right)$
5	Sigmoid (Tanh Hiperbolik)	$K(x,y) = \tanh(\sigma(x.y)+c)$

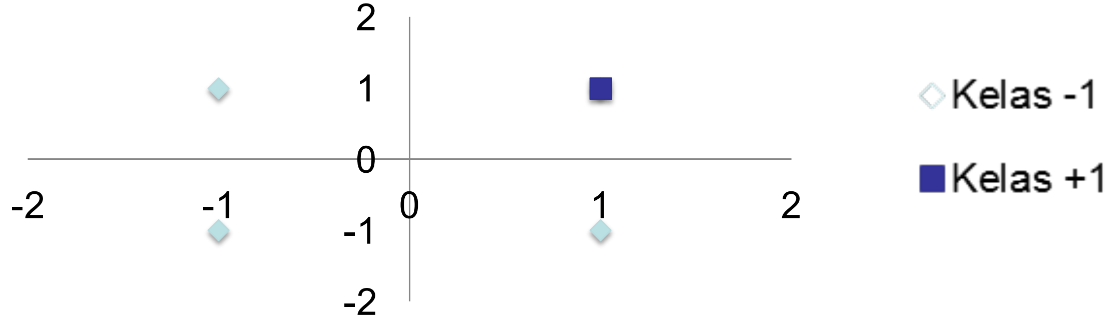
Contoh

- Contoh SVM Linier pada dataset berikut :

  Tentukan Hyperplanenya !

x_1	x_2	Kelas (y)
1	1	1
1	-1	-1
-1	1	-1
-1	-1	-1

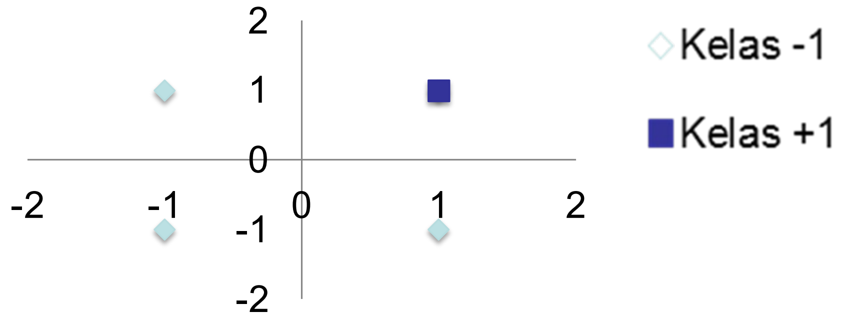
- Bentuk Visualisasi data :



Contoh

- Contoh SVM Linier :

x_1	x_2	Kelas (y)
1	1	1
1	-1	-1
-1	1	-1
-1	-1	-1



- Karena ada dua fitur (x_1 dan x_2), maka w juga akan memiliki 2 fitur (w_1 dan w_2).
- Formulasi yang digunakan adalah sebagai berikut :
  - Meminimalkan nilai : $\frac{1}{2}\|w\|^2 = \frac{1}{2}(w_1{}^2 + w_2{}^2)$
  - Syarat :

$$y_i(w.x_i + b) \geq 1, \qquad i = 1,2,3, \ldots\ldots, N$$

Contoh

- Karena ada dua fitur (x_1 dan x_2), maka w juga akan memiliki 2 fitur (w_1 dan w_2).
- Formulasi yang digunakan adalah sebagai berikut :
  - Meminimalkan nilai margin :

$$\frac{1}{2}\|w\|^2 = \frac{1}{2}(w_1^2 + w_2^2)$$

  - Syarat :

$$y_i(w.x_i + b) \geq 1, dimana\ i = 1,2,3, \ldots, N$$
$$y_i(w_1.x_1 + w_2.x_2 + b) \geq 1$$

x_1	x_2	Kelas (y)
1	1	1
1	-1	-1
-1	1	-1
-1	-1	-1

Sehingga didapatkan beberapa persamaan berikut :

1. $(w_1 + w_2 + b) \geq 1, untuk\ y_1 = 1, x_1 = 1, x_2 = 1$
2. $(-w_1 + w_2 - b) \geq 1, untuk\ y_2 = -1, x_1 = 1, x_2 = -1$
3. $(w_1 - w_2 - b) \geq 1, untuk\ y_3 = -1, x_1 = -1, x_2 = 1$
4. $(w_1 + w_2 - b) \geq 1, untuk\ y_4 = -1, x_1 = -1, x_2 = -1$

Contoh

- Dari 4 persamaan berikut, carilah nilai w_1, w_2 dan b
  1. $(w_1 + w_2 + b) \geq 1$
  2. $(-w_1 + w_2 - b) \geq 1$
  3. $(w_1 - w_2 - b) \geq 1$
  4. $(w_1 + w_2 - b) \geq 1$

Tahap 1: Jumlahkan persamaan (1) dan (2)

$(w_1 + w_2 + b) \geq 1$
$(-w_1 + w_2 - b) \geq 1$
------------------------------ +
$2w_2 = 2$
Maka $w_2 = 1$

Tahap 2: Jumlahkan persamaan (1) dan (3):

$(w_1 + w_2 + b) \geq 1$
$(w_1 - w_2 - b) \geq 1$
------------------------------ +
$2w_1 = 2$
Maka $w_1 = 1$

Tahap 3: Jumlahkan persamaan (2) dan (3):

$(-w_1 + w_2 - b) \geq 1$
$(w_1 - w_2 - b) \geq 1$

$-2b = 2$
Maka $b = -1$

Tahap 4: didapatkan persamaan *hyperplane*

$w_1.x_1 + w_2.x_2 + b = 0$

$x_1 + x_2 + 1 = 0$

$x_2 = 1 - x_1$

Contoh

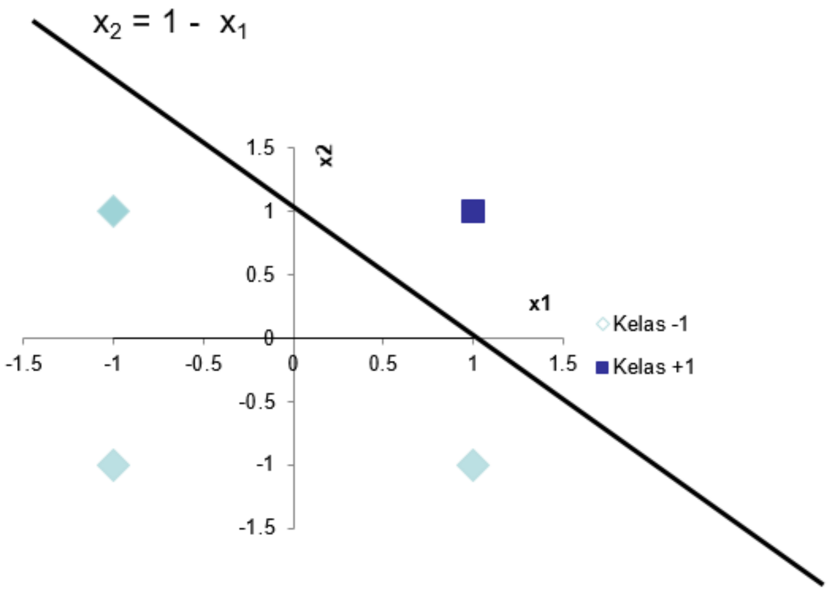
Visualisası garis hyperplane (sebagai fungsi klasifikasi) :

$w_1x_1 + w_2x_2 + b = 0$

$x_1 + x_2 - 1 = 0$

$x_2 = 1 - x_1$

x_1	$x_2 = 1 - x_1$
-2	3
-1	2
0	1
1	0
2	-1



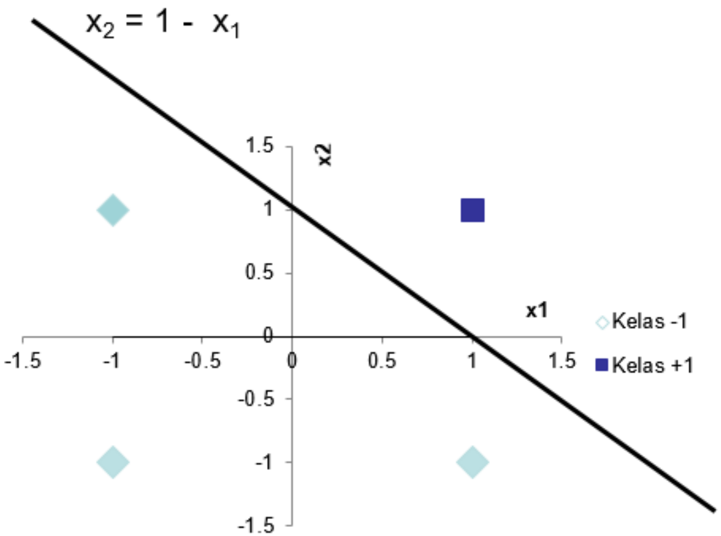
Contoh

Misalkan diketahui data uji/ data testing berikut :

Diketahui : $f(x) = x_1 + x_2 - 1$

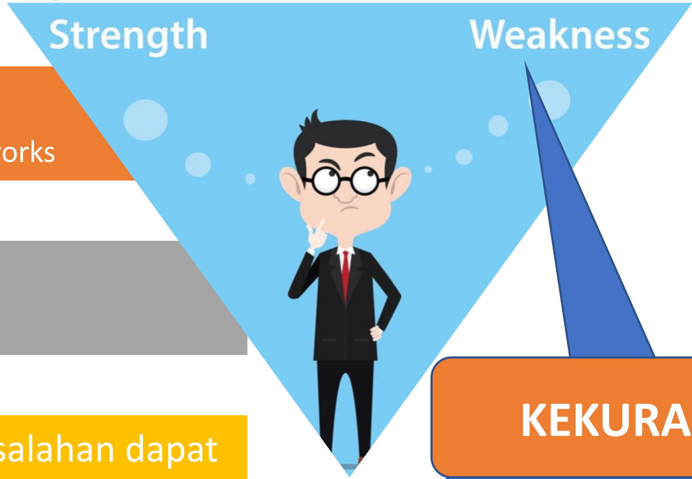
Kelas = sign(f(x))

No	Data Uji		Hasil Klasifikasi
	x_1	x_2	Kelas = sign($x_1 + x_2 - 1$)
1	1	5	sign (1 + 5 - 1) = +1
2	-1	4	sign (-1 + 4 - 1) = +1
3	0	7	sign (0 + 7 - 1) = +1
4	-9	0	sign (-9 + 0 - 1) = -1
5	2	-2	sign (2 - 2 - 1) = -1



Kelebihan dan Kekurangan

- Training relatif mudah
  - tidak mengalami local optimal, tidak seperti pada neural networks
- Relatif baik untuk data berdimensi tinggi
- Tradeoff antara kompleksitas classifier dan kesalahan dapat dikontrol secara eksplisit
- Data non-tradisional seperti string dan tree dapat digunakan sebagai input ke SVM, bukan vektor fitur



KEKURANGAN

Membutuhkan fungsi kernel yang bagus

SVM vs Neural Networks

SVM	ANN
Konsep yang relatif baru	Men-generalisasi dengan baik tetapi tidak memiliki dasar matematika
Sifat generalisasi yang bagus	Dapat dengan mudah dipelajari secara bertahap
Sulit untuk dipelajari → teknik QP (Quadratic Programming)	Untuk mempelajari fungsi kompleks → gunakan struktur multi layer yang kompleks.
Menggunakan kernel, dapat mempelajari fungsi yang sangat kompleks	

- TERIMA KASIH -

www.its.ac.id its_campus institut teknologi sepuluh nopember