Support Vector Machines

IF-3270 Pembelajaran Mesin

Teknik Informatika ITB

Modul 5: Support Vector Machine

01 SVM: What & Why?

IF3270 - Pembelajaran Mesin
(Machine Learning)

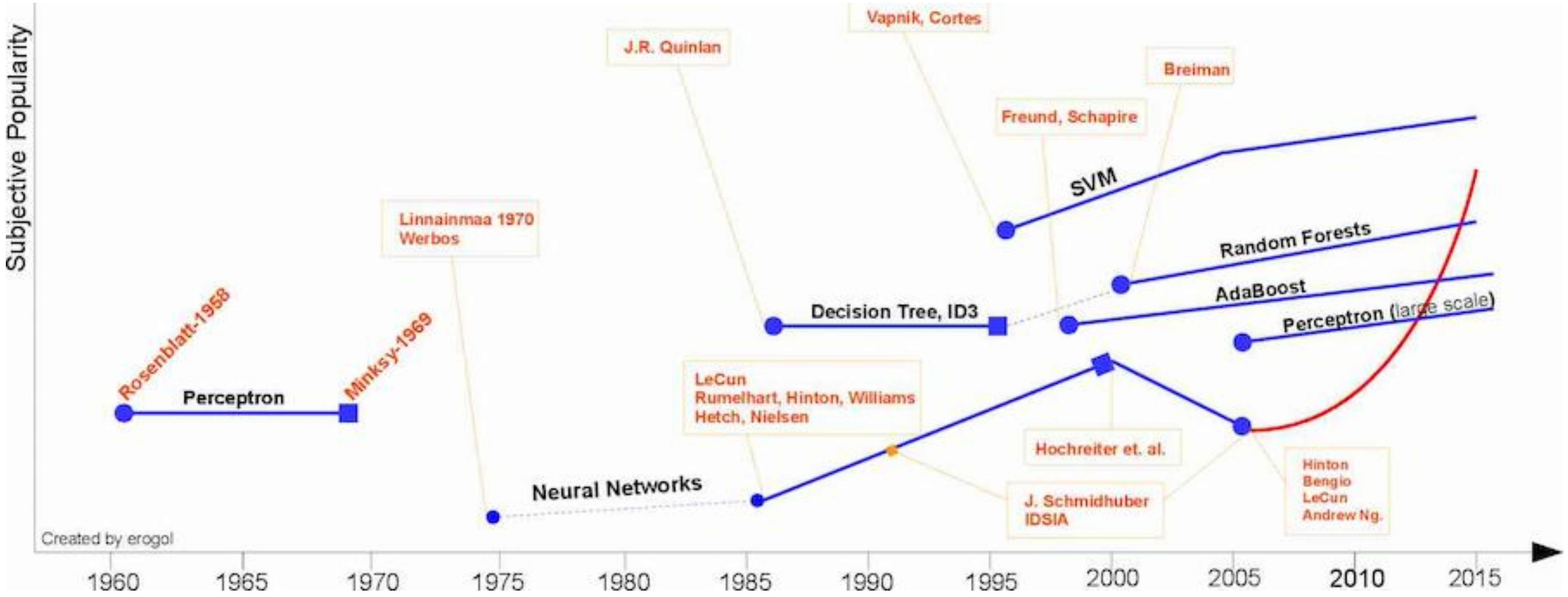
Fariska Z. Ruskanda, S.T., M.T.
(fariska@informatika.org)

KK IF -Teknik Informatika - STEI ITB

Outline

- Sejarah SVM
- Tujuan SVM
- Bidang Pemisah Terbaik
- Klasifikasi Biner – Linear Separability
- Hyperplane Classifier

Support Vector Machine



Eren Golge: https://www.kdnuggets.com/2014/10/deep-learning-make-machine-learning-algorithms-obsolete.html

- SVM diperkenalkan tahun 1992 oleh Vapnik, Boser, & Guyon
- Kinerja baik di berbagai aplikasi seperti *bioinformatics*, klasifikasi teks, pengenalan tulisan tangan dan lain-lain.

SVM

- 1980an
  - DTL dan NN memungkinkan pembelajaran nonlinear yang efisien
  - Kurang didukung dasar teoritis dan memungkinkan terjadinya local minima
- 1990an
  - Algoritma pembelajaran yang efisien untuk fungsi non linier berbasis teori komputasi

SVM Introduction

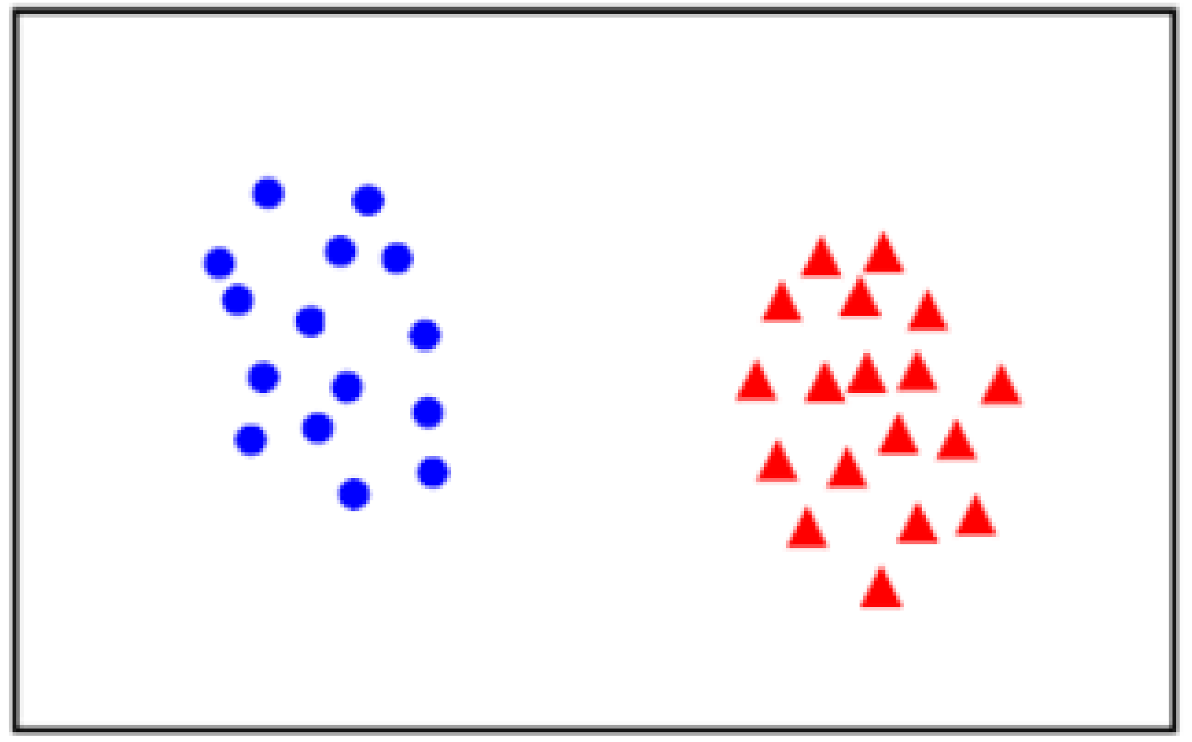
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152). ACM.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

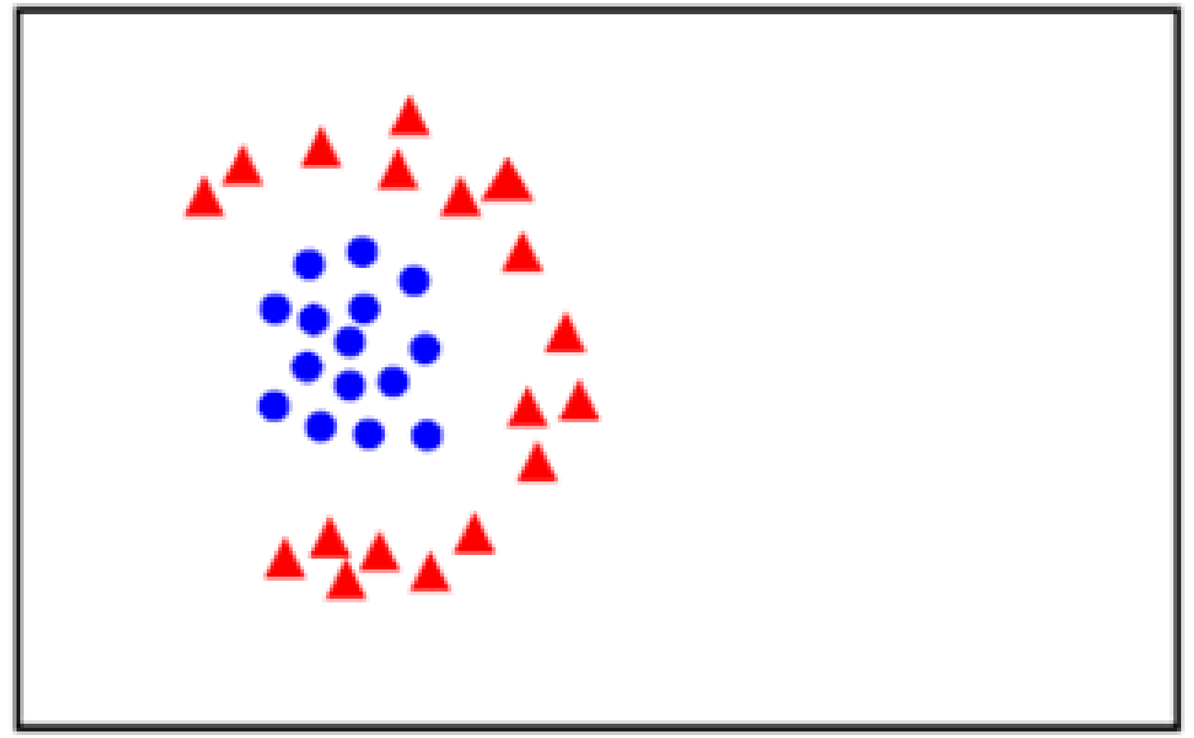
Klasifikasi Biner

Given training data $(\mathbf{x}_i, y_i)$ for $i = 1 \dots N$, with $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$, learn a classifier $f(\mathbf{x})$ such that

$$f(\mathbf{x}_i) \begin{cases} \geq 0 & y_i = +1 \\ < 0 & y_i = -1 \end{cases}$$

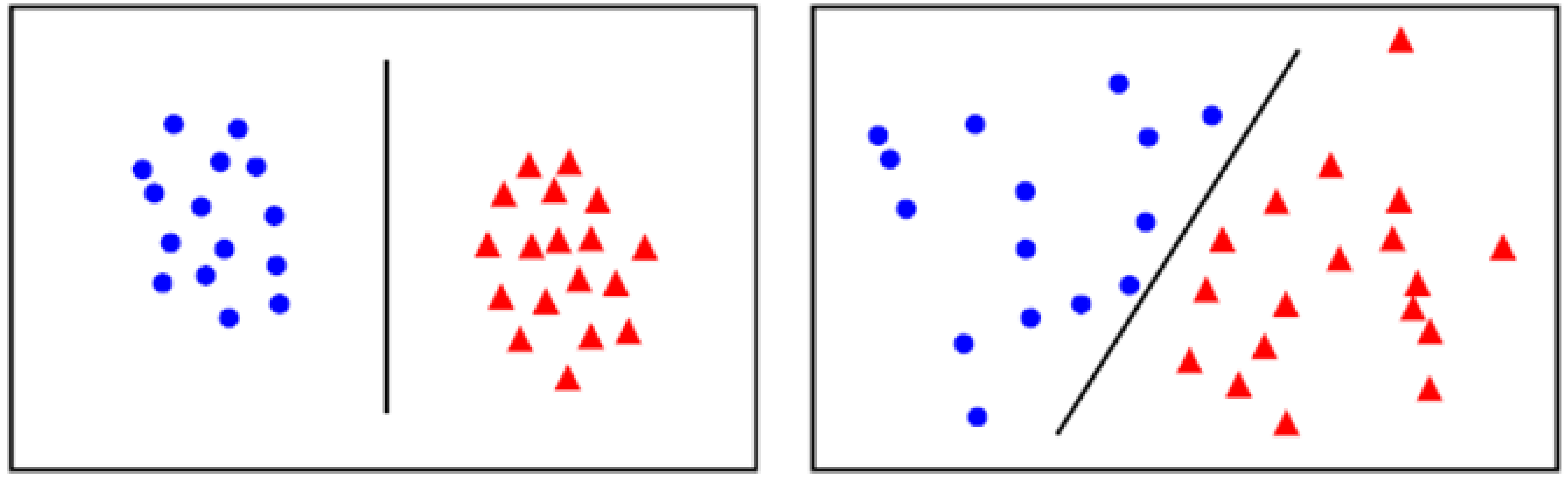
i.e. $y_i f(\mathbf{x}_i) > 0$ for a correct classification.



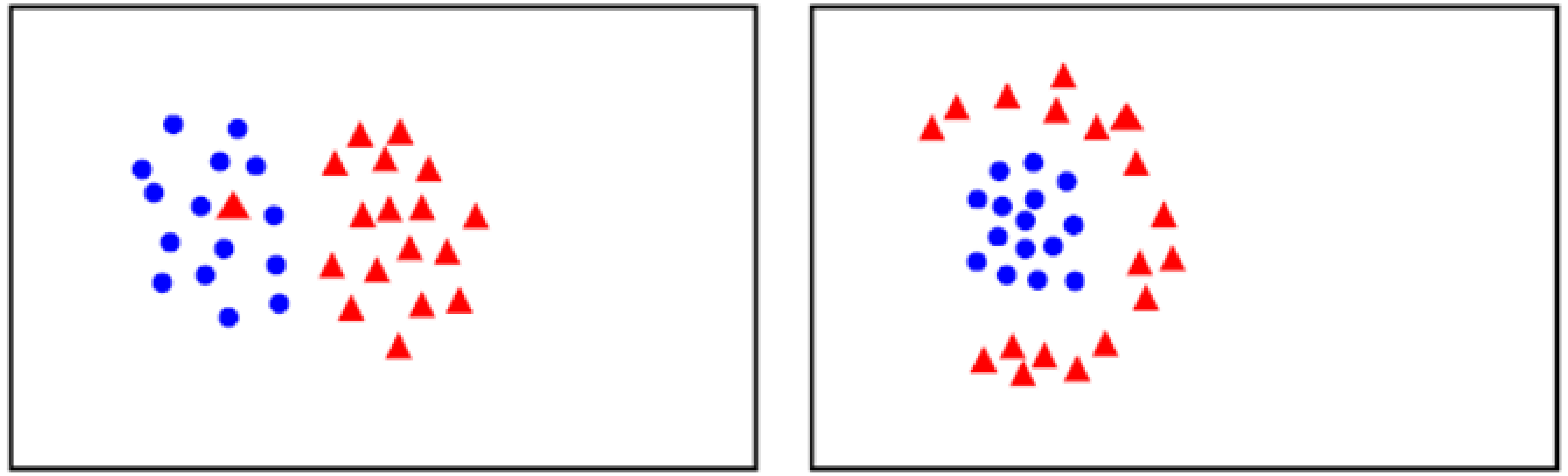


Linear Separability

linearly separable



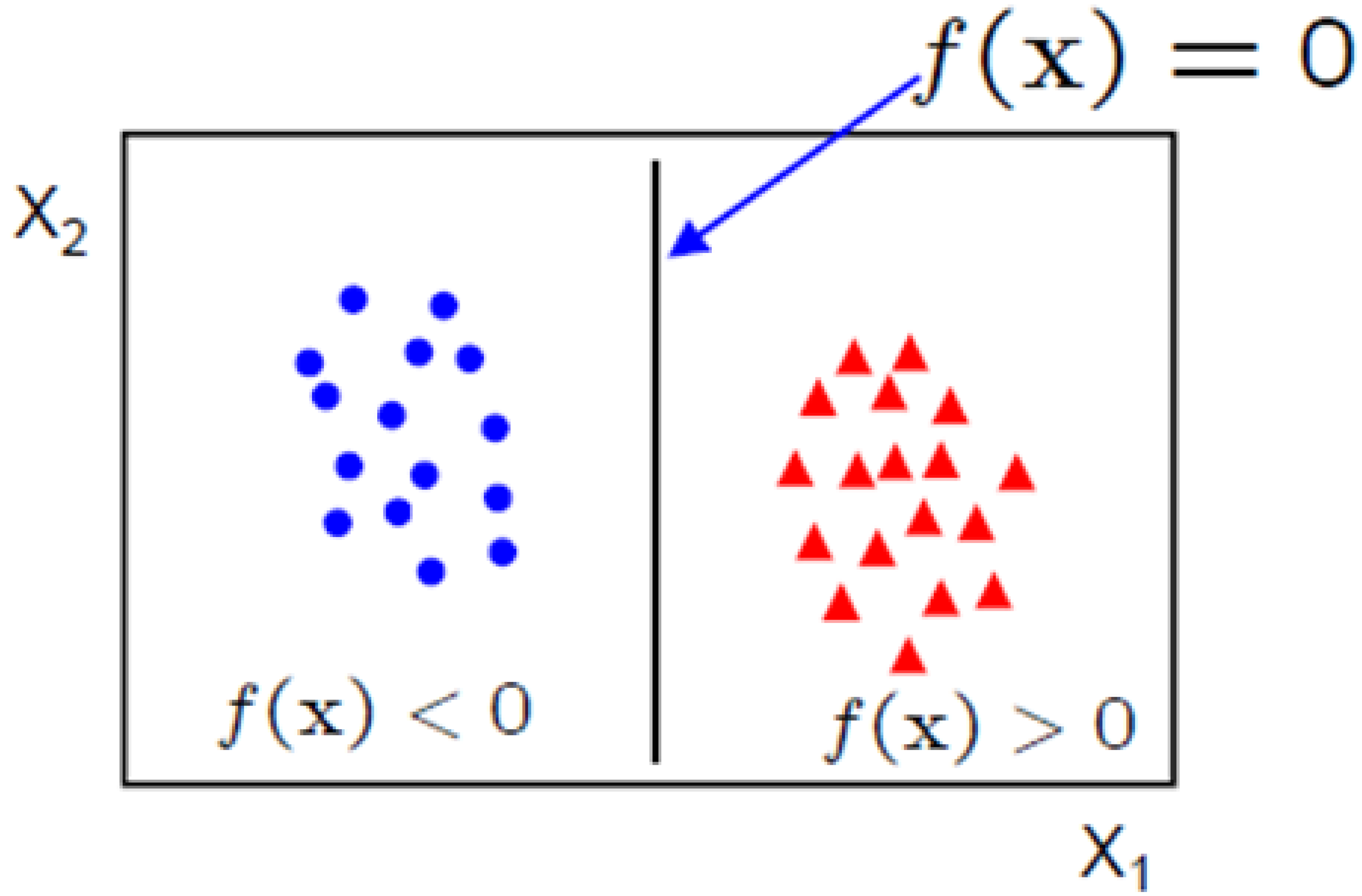
not linearly separable



Linear Classifier

A linear classifier has the form

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$$

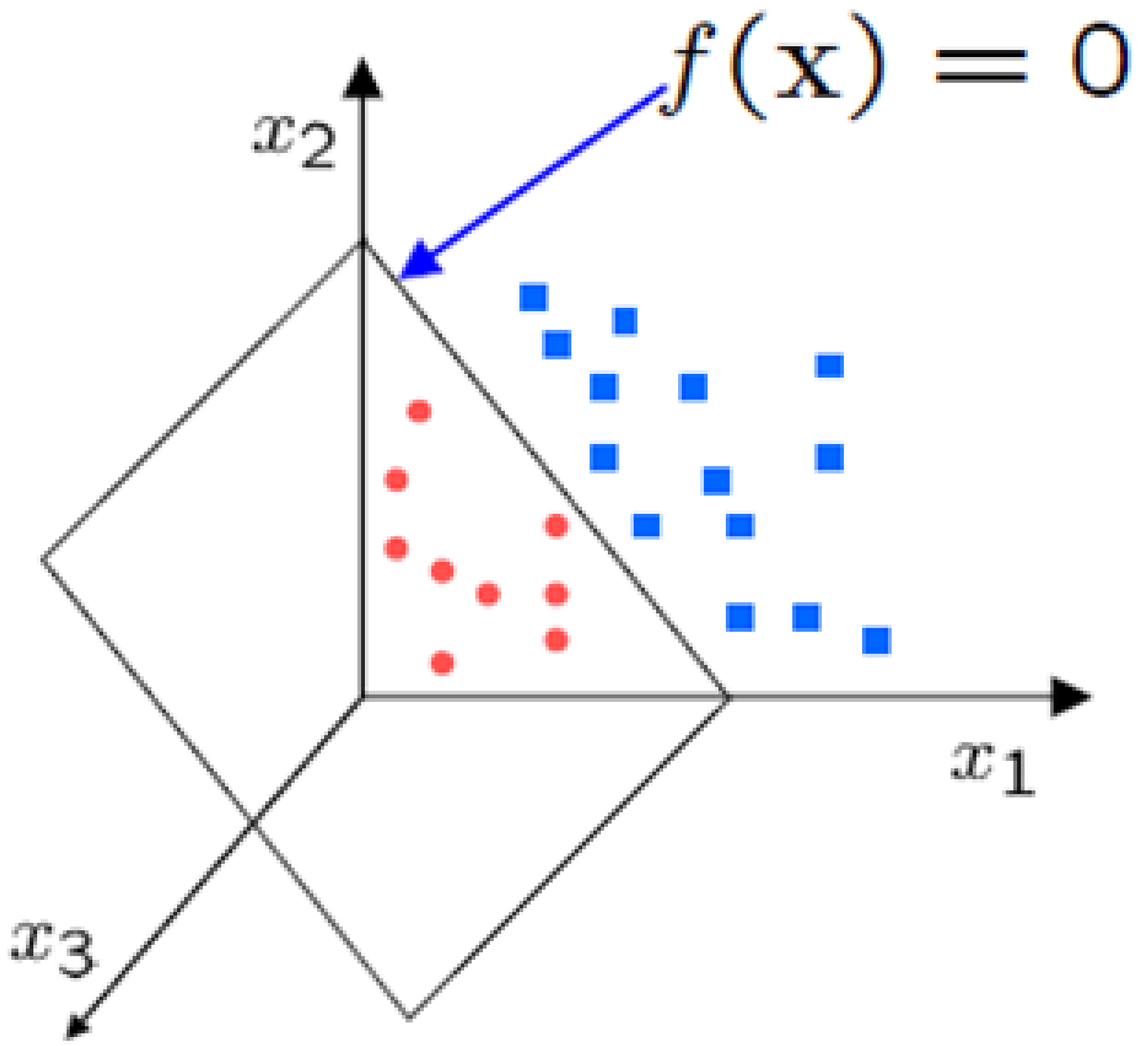


- in 2D the discriminant is a line
- $\mathbf{w}$ is the normal to the line, and b the bias
- $\mathbf{w}$ is known as the weight vector

Linear Classifier

A linear classifier has the form

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$$



- in 3D the discriminant is a plane, and in nD it is a hyperplane

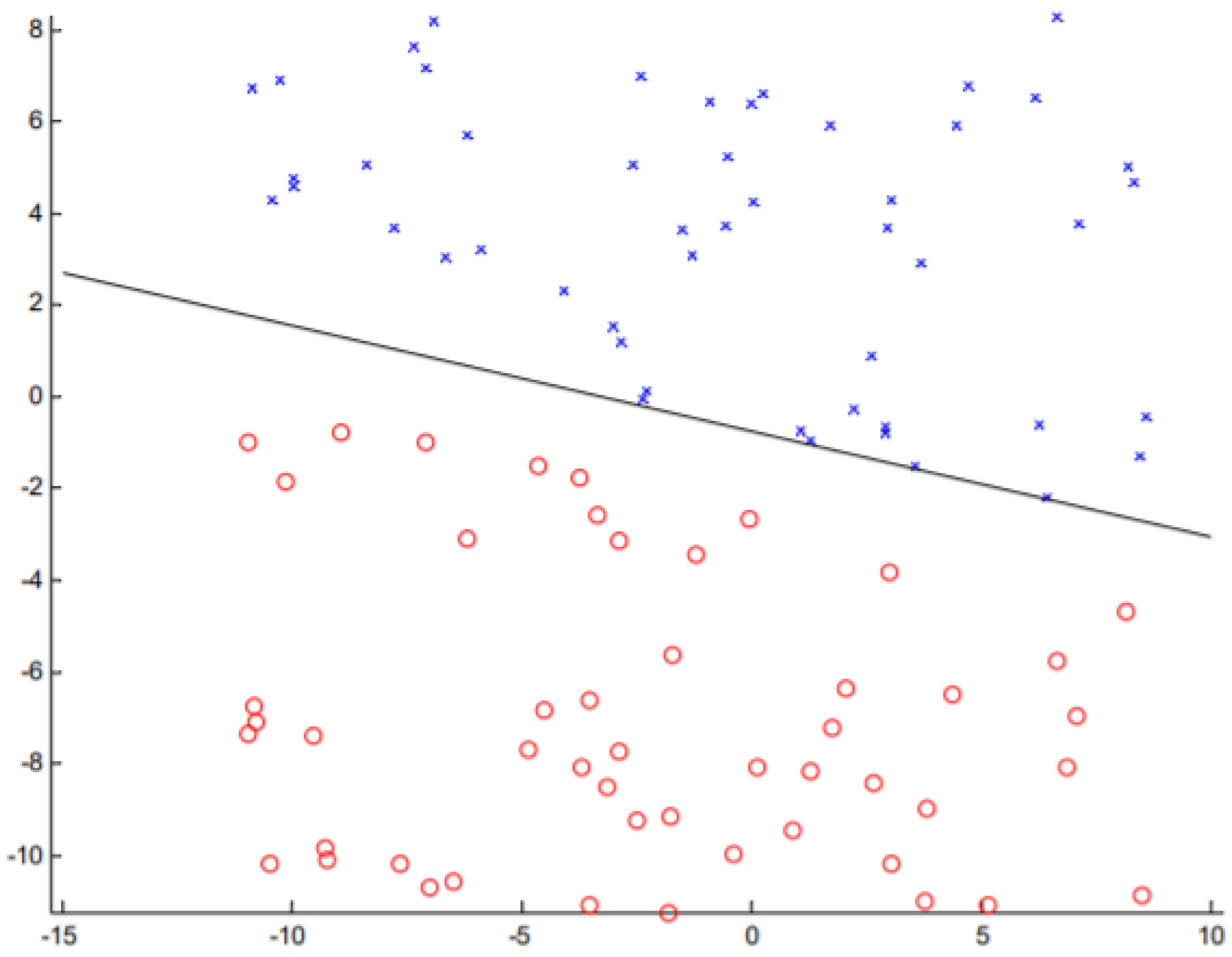
For a K-NN classifier it was necessary to `carry' the training data

For a linear classifier, the training data is used to learn **w** and then discarded

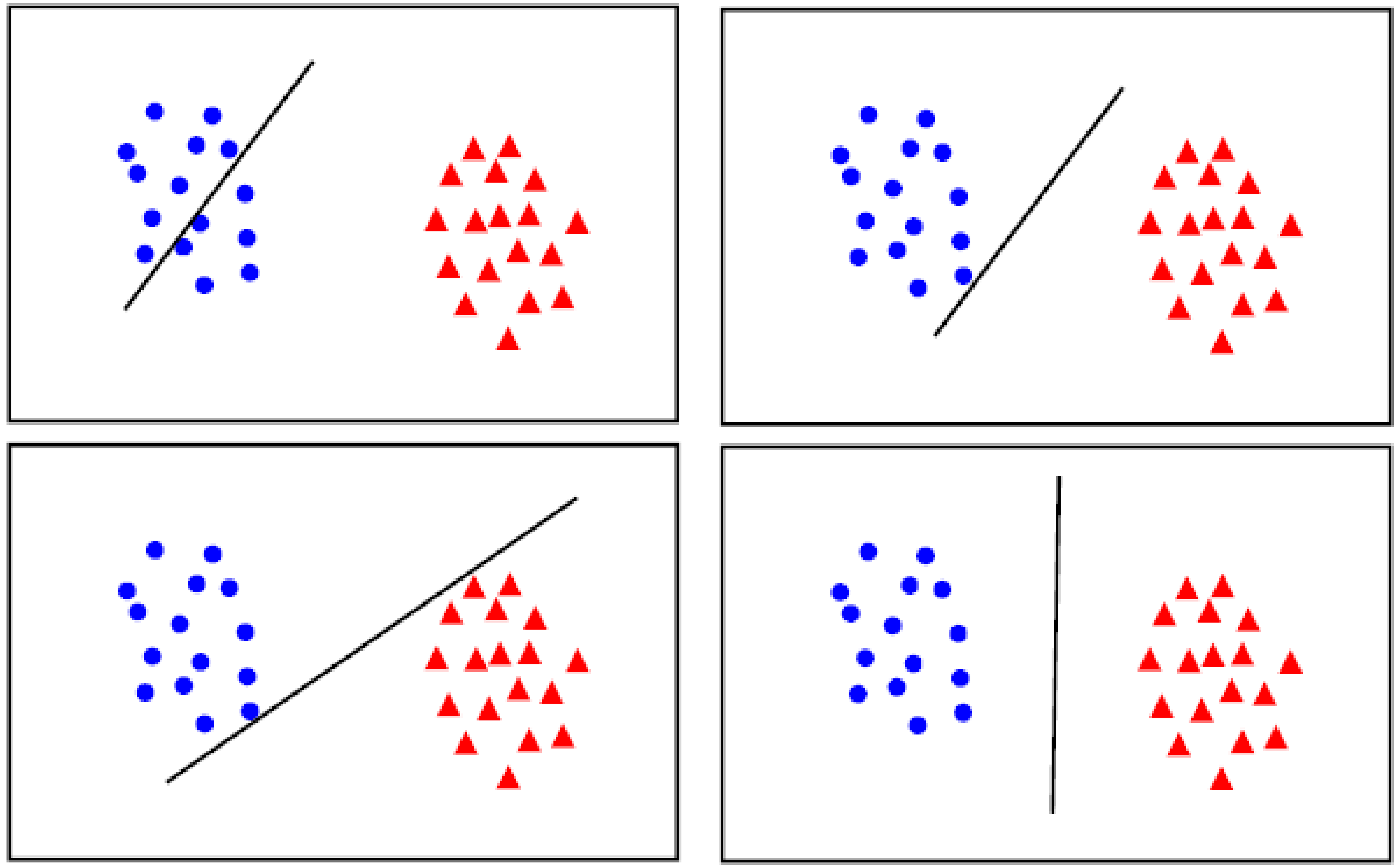
Only **w** is needed for classifying new data

Perceptron Weakness

- Perceptron biggest weakness is that it will not find the same hyperplane every time.
  - Not all separating hyperplanes are equals.
  - If the Perceptron gives you a hyperplane that is very close to all the data points from one class, you have a right to believe that it will generalize poorly when given new data.
- After an accurate hyperplane is found, the training process will stop and it is considered to have converged.



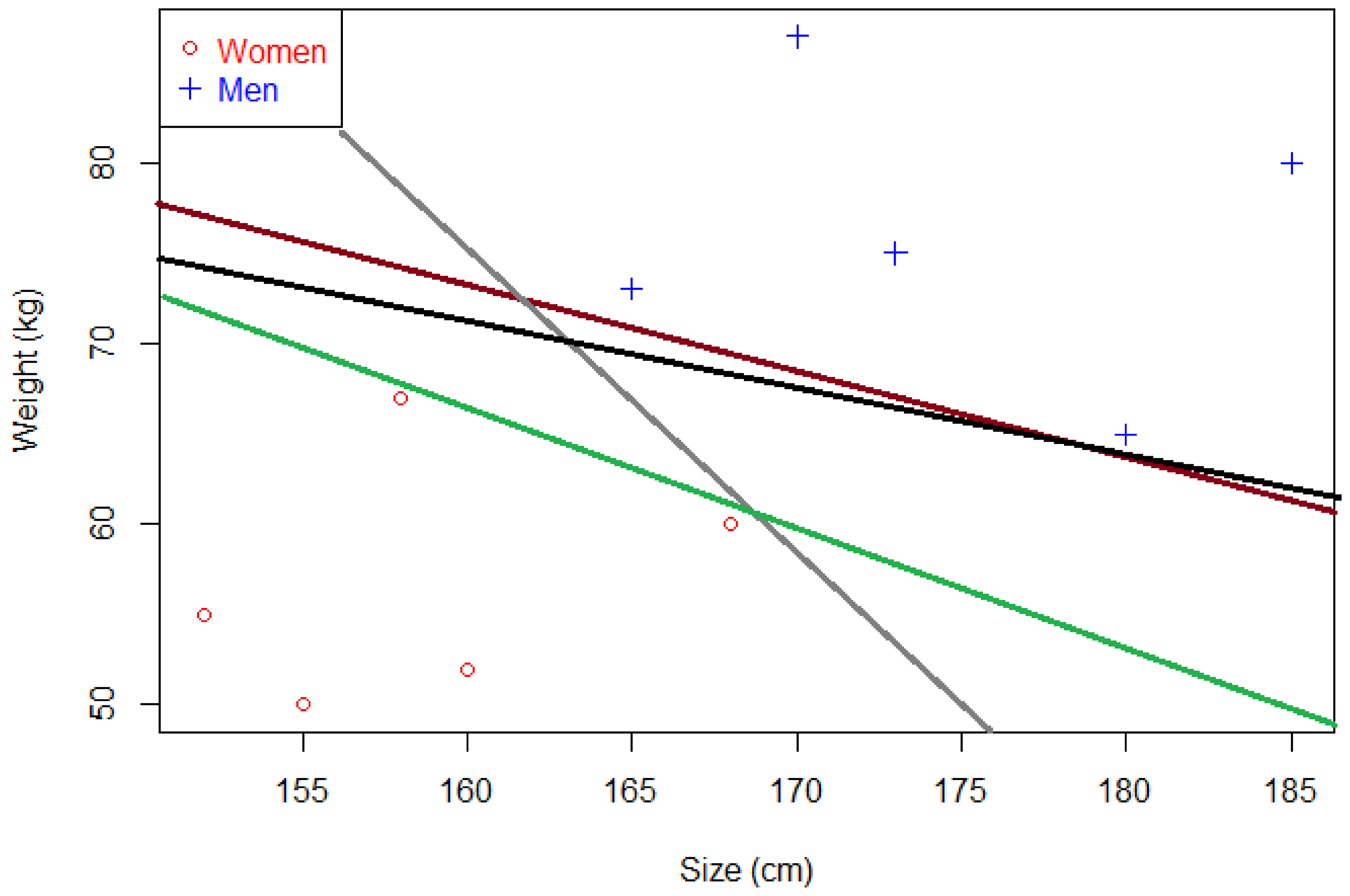
Bidang Pemisah Terbaik?

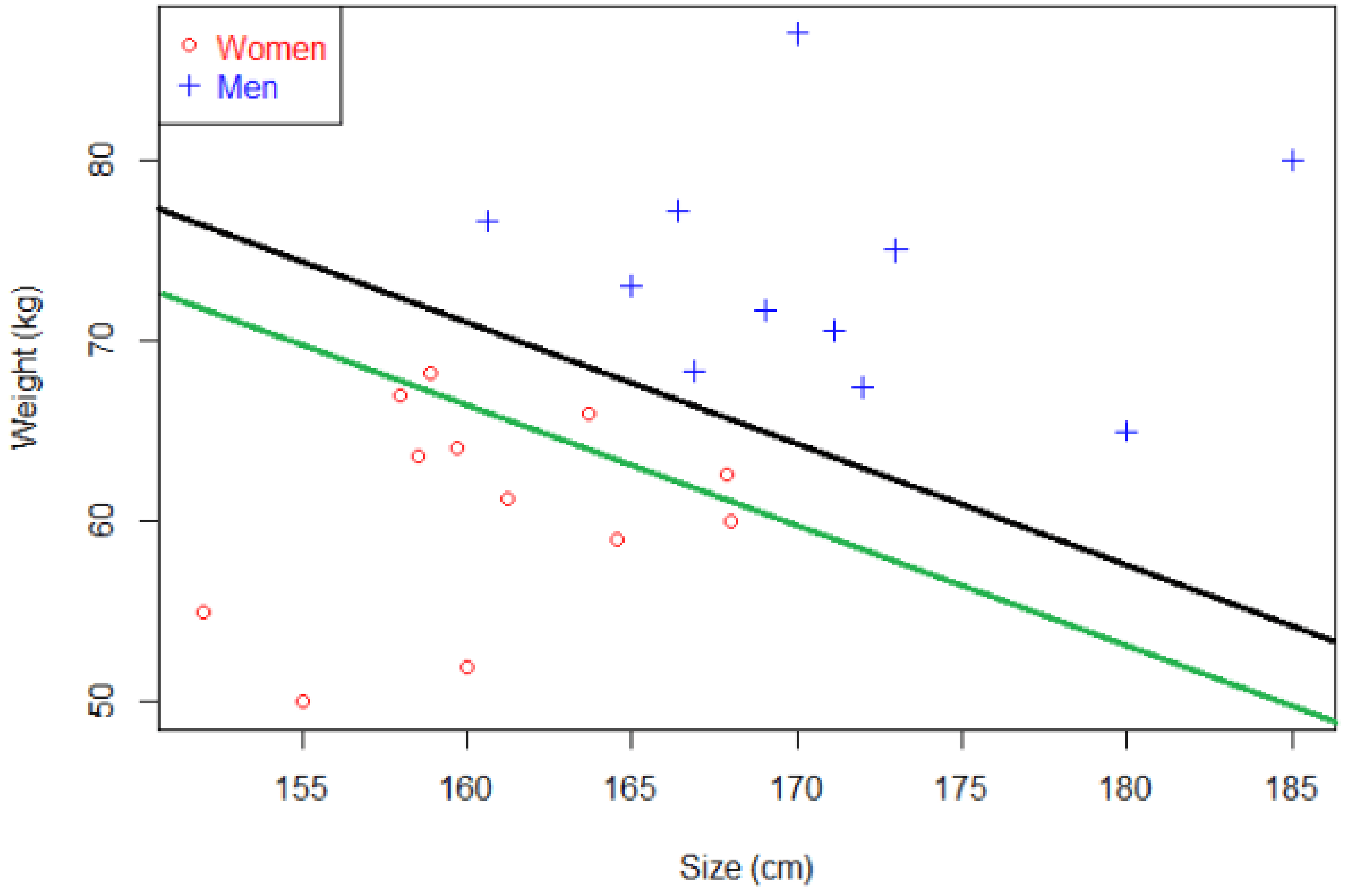


- Mengapa?

https://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf

Bidang Pemisah Terbaik (lanj)





Kiri: semua bidang pemisah valid karena memisahkan kedua kelas pada training data.
Kanan: real-life data. Bidang pemisah hitam lebih baik daripada hijau.

SVM Objective

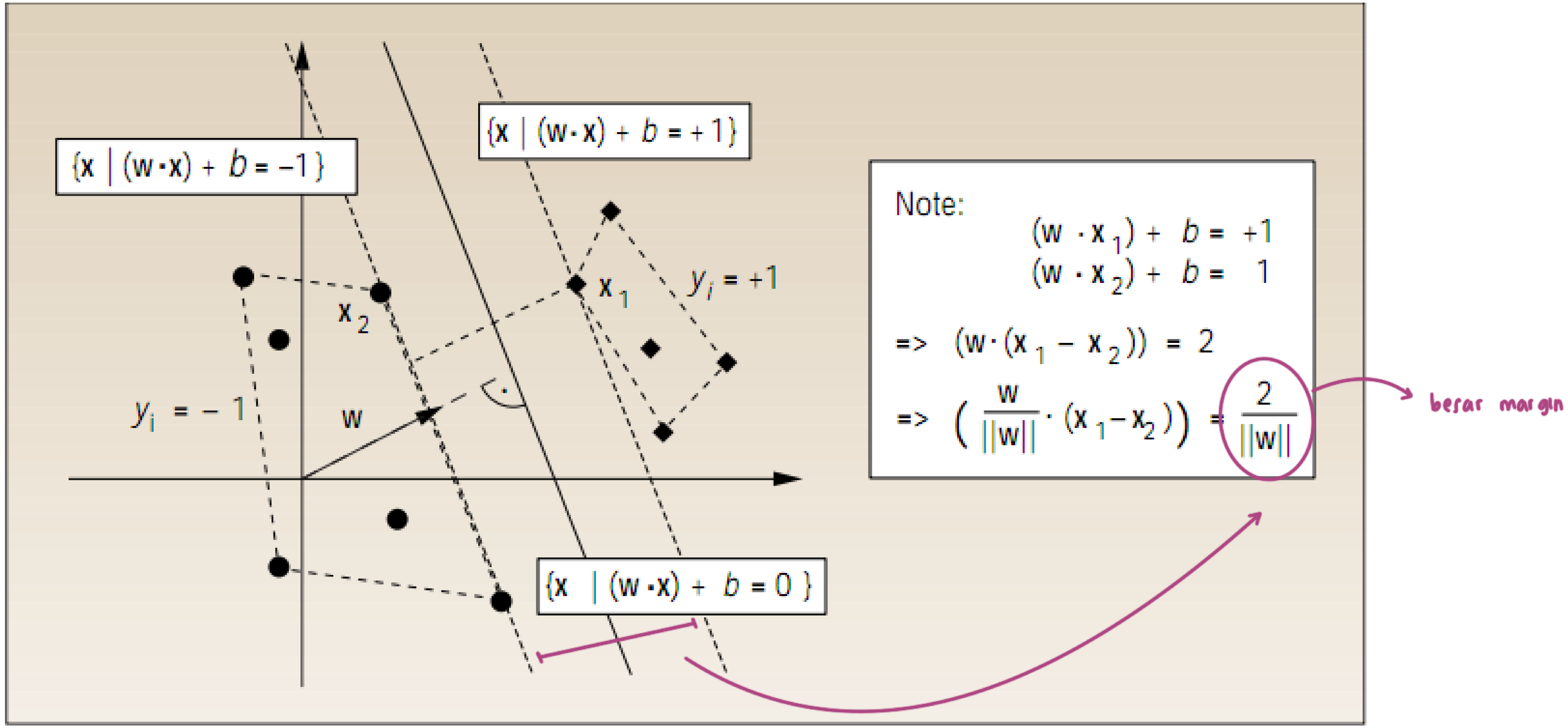
- Objective of the SVM **is to find the optimal separating hyperplane which maximizes the margin of the training data.** There will never be any data point inside the margin. → tdk boleh ada data di dlm margin
- Menggunakan optimasi kuadratik untuk menghindari 'local minimum' isu yang ada pada NN (Greedy)
- Menggunakan fungsi kernel untuk memisahkan non-linear region

https://www.svm-tutorial.com/2015/06/svm-understanding-math-part-3/

Hyperplane Classifier

- Hipotesis:
- x1, x2 ∈ training data

$$f(x) = sign(\vec{w}.\vec{x} + b); \vec{w}, \vec{x} \in \Re^N; b \in \Re$$



Vector Direction

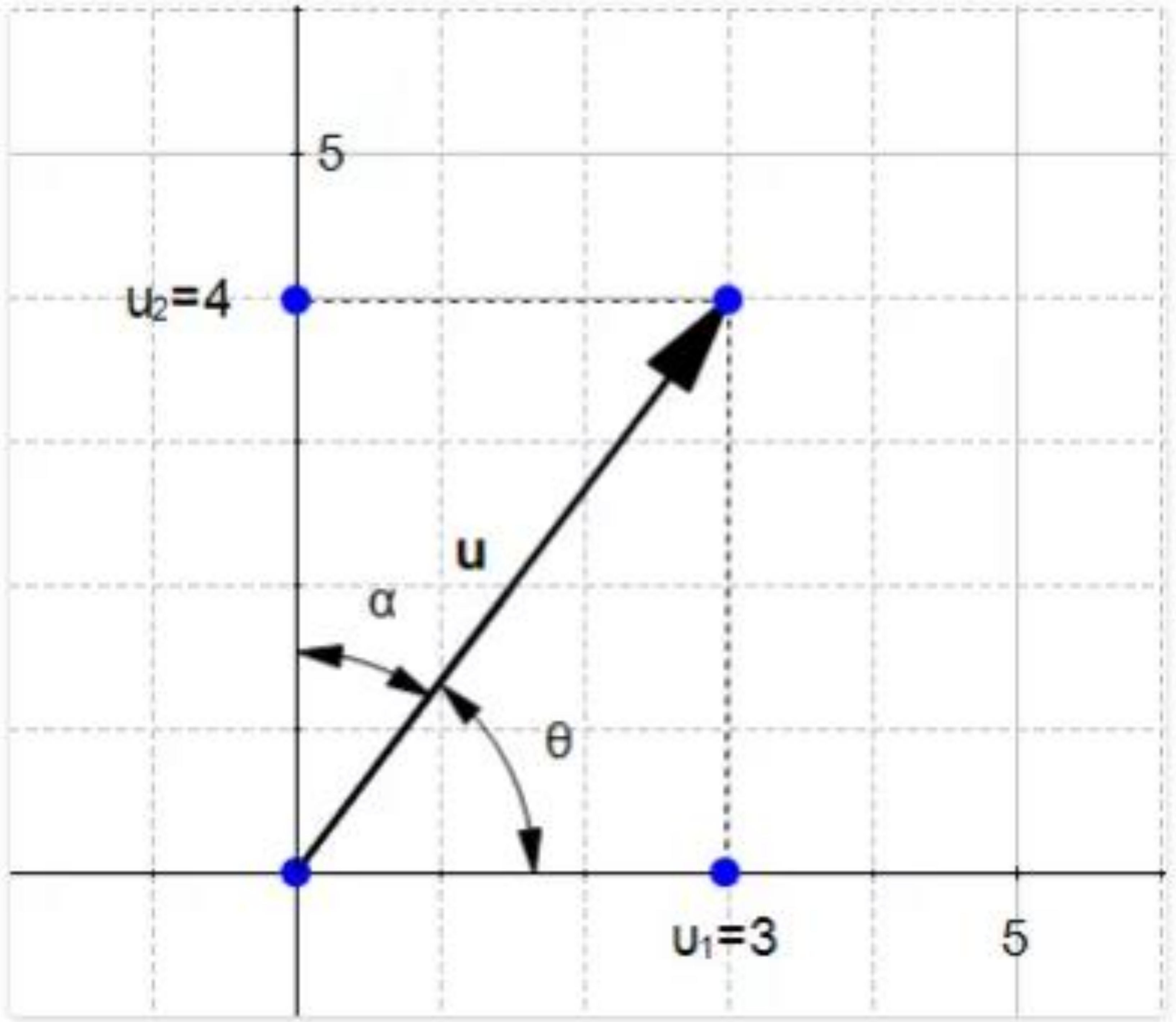


Figure 4 - direction of a vector

u(u1,u2) with u1=3 and u2=4

$$cos(\theta) = \frac{u_1}{\|u\|} \qquad cos(\alpha) = \frac{u_2}{\|u\|}$$

- Naive definition 1: The direction of the vector u is defined by the angle θ with respect to the horizontal axis, and with the angle α with respect to the vertical axis.
- Naive definition 2: The direction of the vector u is defined by the cosine of the angle θ and the cosine of the angle α.

02 SVM for Linearly Separable Data

IF3270 Pembelajaran Mesin