From ML-A to ML-B

Margin-based Linear Classification

Kernels – Linear Classification of non-linearly transformed data

Nirupam Gupta
(With minor revision by Yevgeny Seldin)

Recall: Linear Classifier

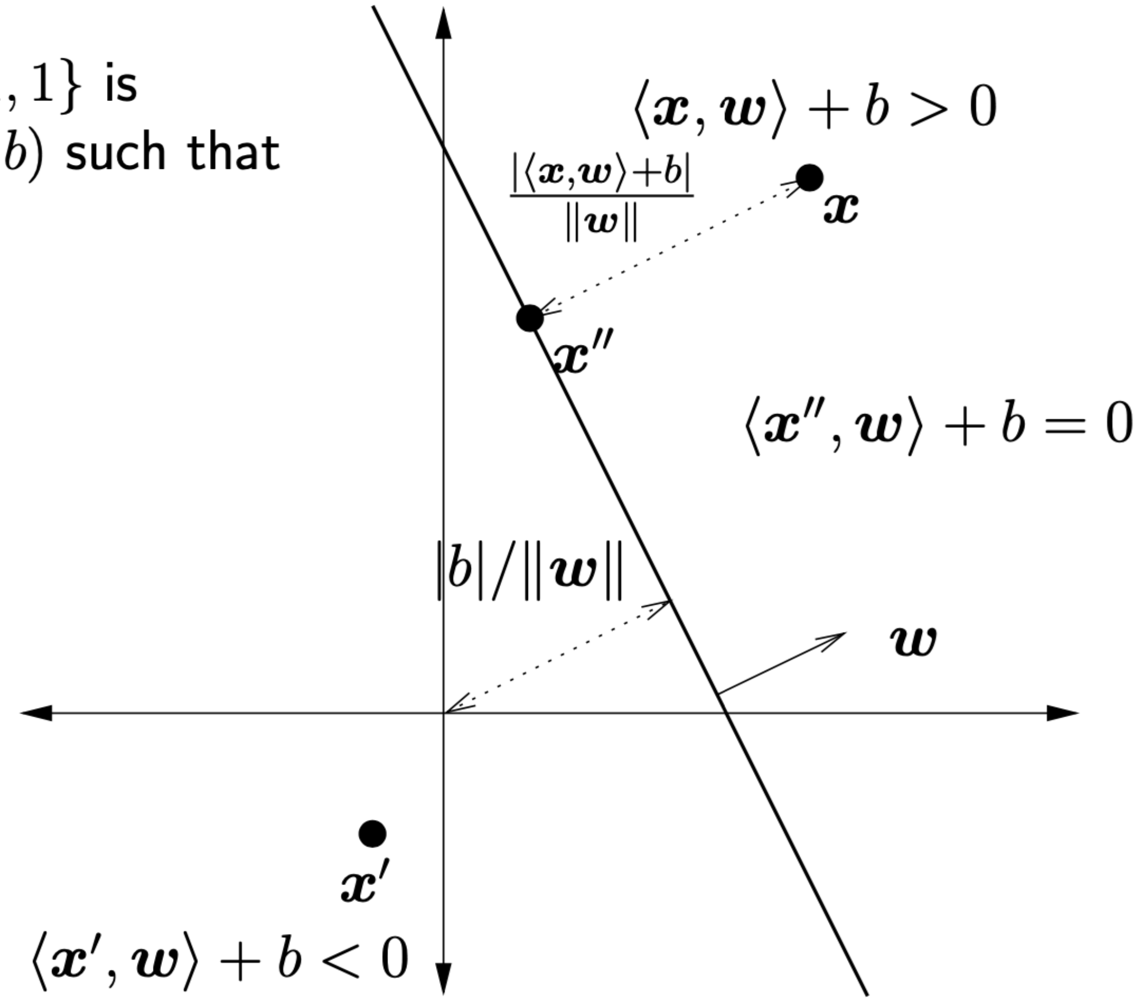
$S = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}$, $\boldsymbol{x}_i \in \mathbb{R}^d$, $y_i \in \{-1, 1\}$ is linearly separable if there exists a hyperplane $(\boldsymbol{w}, b)$ such that for all $i = 1, \ldots, N$

$$y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) > 0$$

which implies

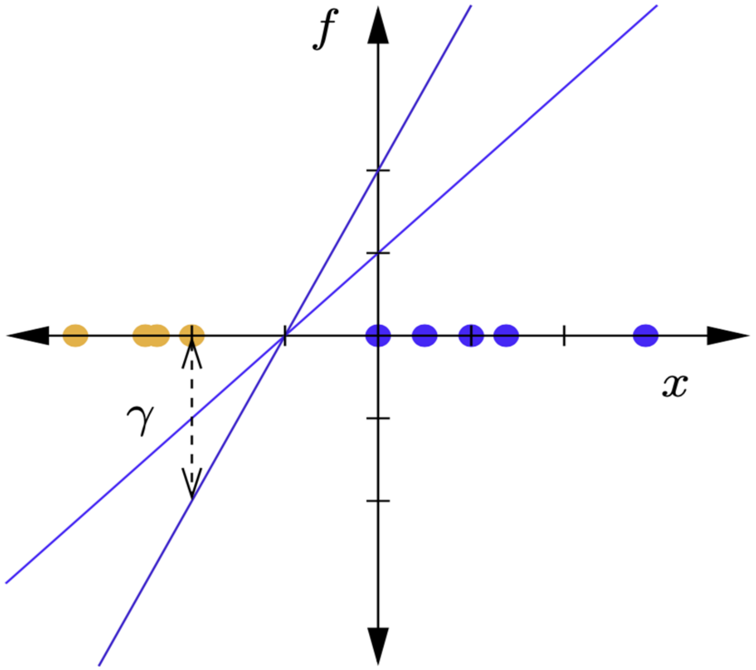
$$y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) \geq \gamma$$

for some $\gamma > 0$.

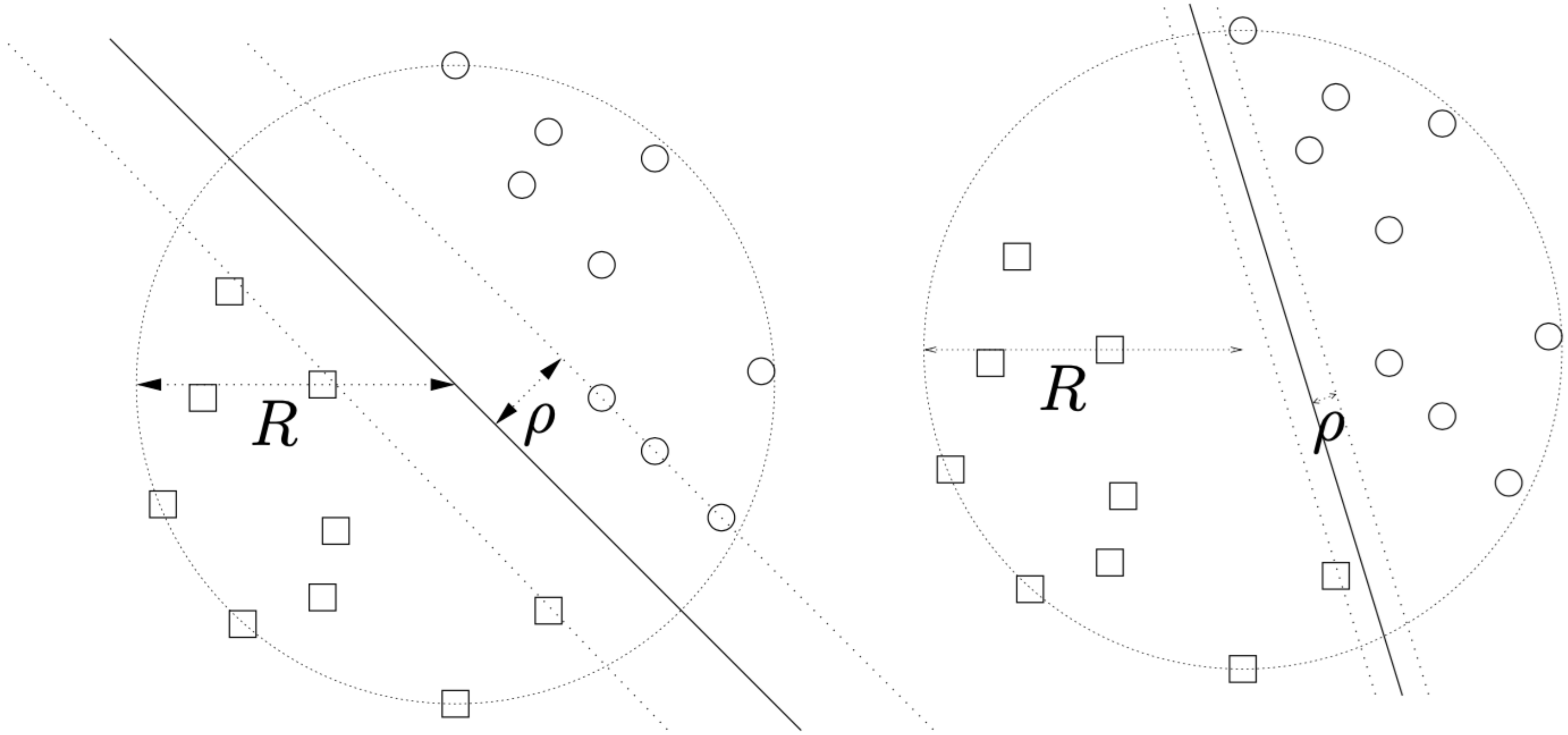


Inherent Degrees of Freedom

- (cw, cb) yields the same separating hyperplane for any $c > 0$
- "Small" tilts of the hyperplane yield identical labeling (identical empirical error)



How to Determine the Hyperplane?



Support Vector Machine (SVM): Maximizing the margin

Linear SVM

Given linearly separable training data $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}$, we get rid of the inherent degree of freedom in

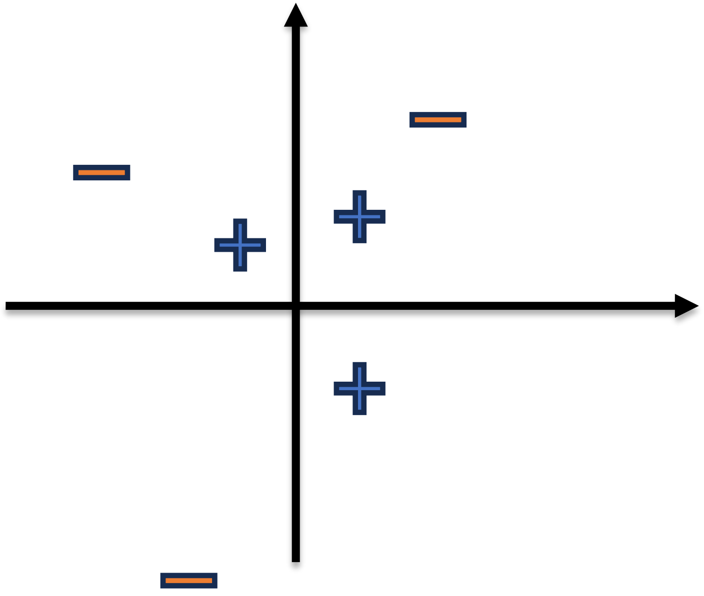
$$\begin{aligned} \text{maximize}_{\boldsymbol{w},b} \quad & \rho = \gamma / \|\boldsymbol{w}\| \\ \text{subject to} \quad & y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) \geq \gamma \ , \quad i = 1, \ldots, N \end{aligned}$$

by fixing $\gamma = 1$ (alternatively $\|\boldsymbol{w}\| = 1$)

And then maximize the margin $1/||w||$

We will learn about how to solve the above problem.

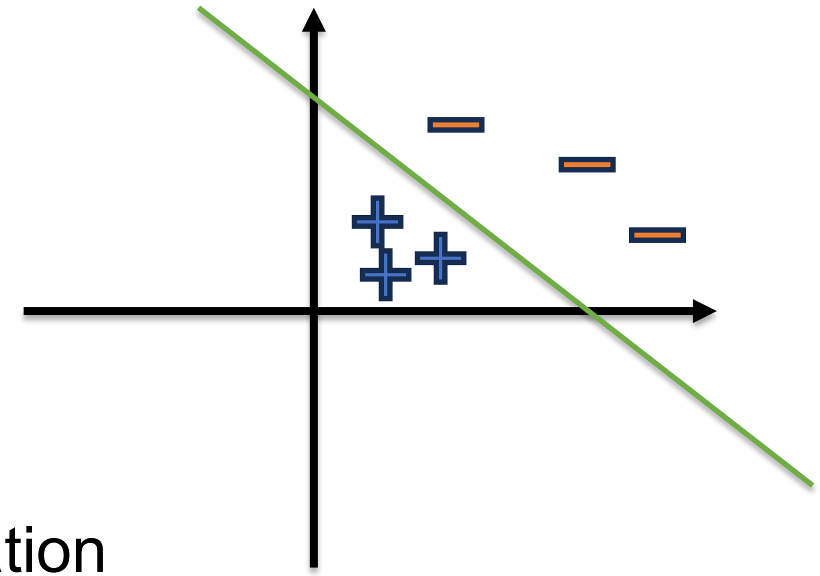
Non-Linear Classifiers (Introduction to Kernels)



$$(x_1, x_2) \rightarrow (z_1, z_2)$$

$$z_1 = x_1^2$$
$$z_2 = x_2^2$$

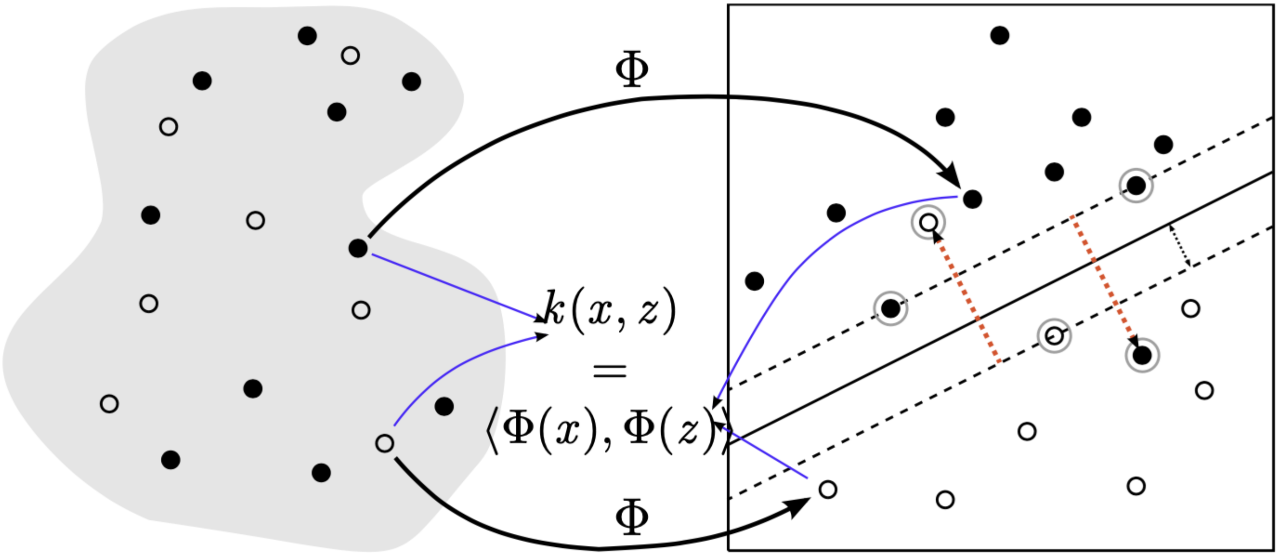
Non-linear transformation



Not linearly separable

Linearly separable!

Non-Linear Classifiers (Introduction to Kernels)



Idea: make learning easier by changing representation

$\mathcal{X}$: input space $\rightarrow \mathcal{Z}$: feature space

Φ: $\mathcal{X} \rightarrow \mathcal{Z}$ (feature map)

and do the classification / regression in $\mathcal{Z}$