Basics of Machine Learning

Part 2

Vera Shalaeva*, 2020

* With use of slides from the course of Introduction to Neural Networks of Pascal Germain

Data and learning model

Training set

A learner observes a finite set $S = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$

Regression

$$\mathbf{x}_i \in \mathbb{R}^d, y_i \in \mathbb{R} \quad f_{\mathbf{w},b} = \mathbf{w} \cdot \mathbf{x} + b$$

Binary classification

$$\mathbf{x}_i \in \mathbb{R}^d, y_i \in \{-1,1\} \quad \text{or} \quad y_i \in \{0,1\}$$

$$f_{\mathbf{w},b} = \mathrm{sgn}[\mathbf{w} \cdot \mathbf{x} + \mathrm{b}]$$

$$f_{\mathbf{w},b} = +1 \text{ if } \mathbf{w} \cdot \mathbf{x} + \mathrm{b} > 0$$

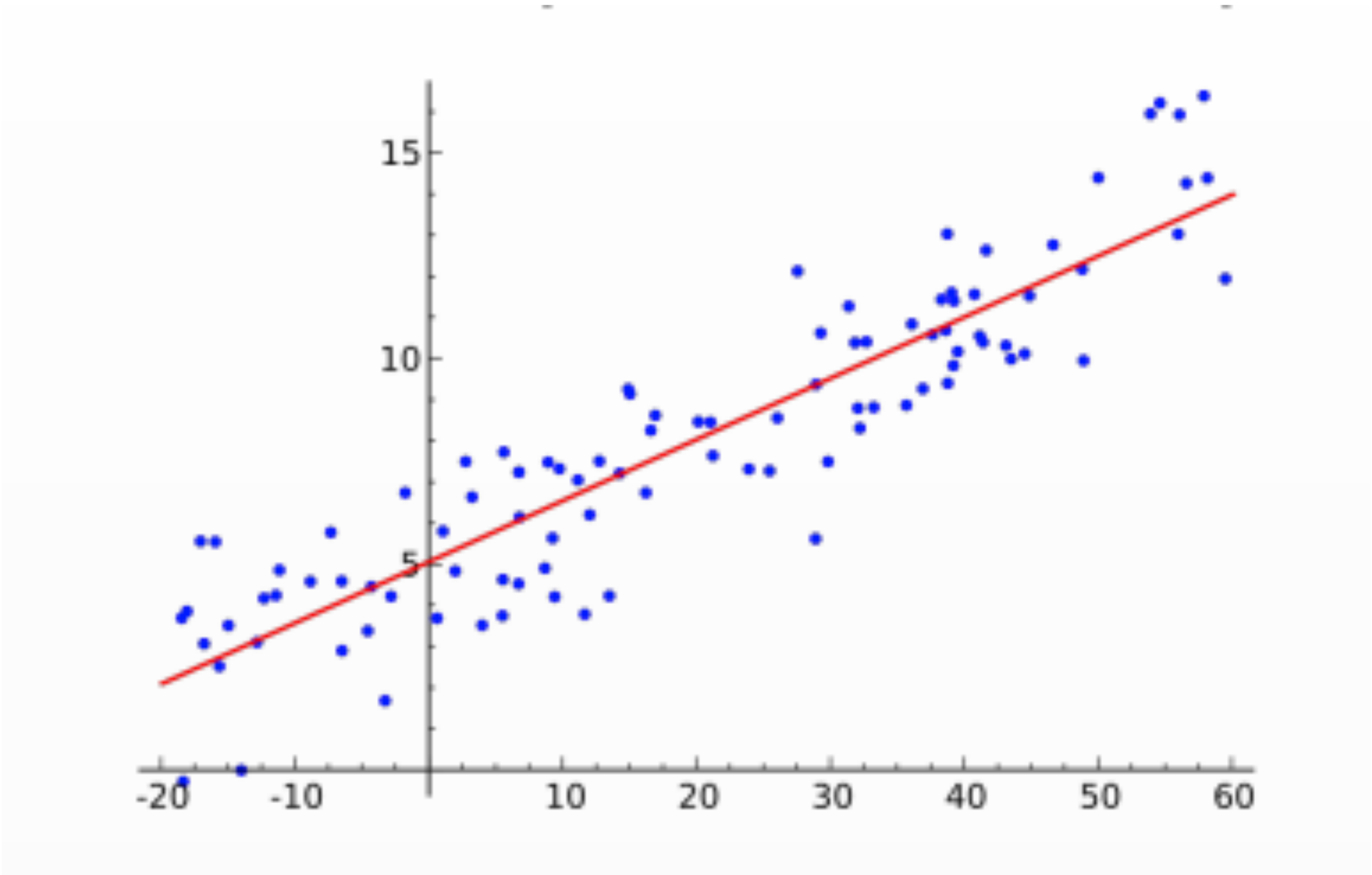
$$f_{\mathbf{w},b} = -1 \text{ otherwise}$$

Note: $f_{\mathbf{w},b} = f_{\mathbf{w}',b'}$ if $\mathbf{w}' = c\mathbf{w}, b' = cb, \forall c > 0$

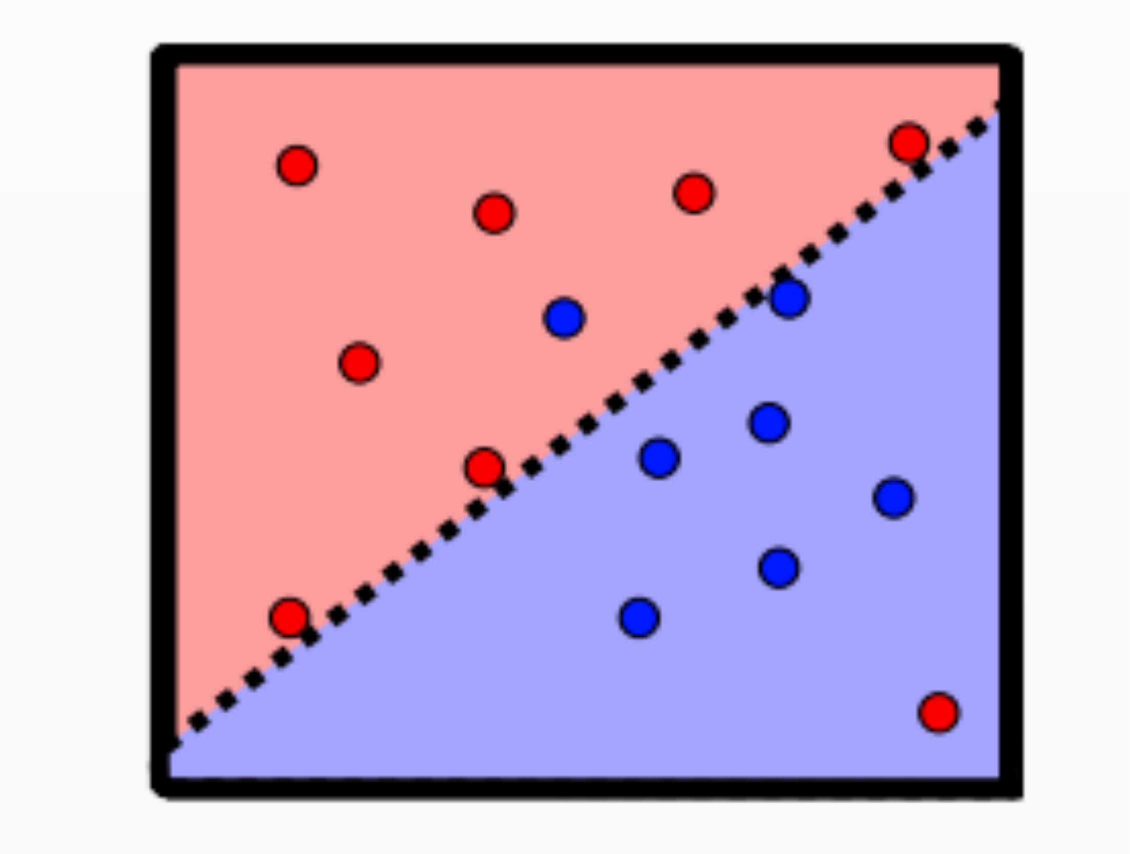
Data and learning model

Regression

We learned linear regression —> predicting real values.



Binary classification



Support Vector Machines (SVM)

Binary classification

$$\mathbf{x}_i \in \mathbb{R}^d, y_i \in \{-1,1\} \quad \text{or} \quad y_i \in \{0,1\}$$

$$f_{\mathbf{w},b} = \text{sgn}[\mathbf{w} \cdot \mathbf{x} + \text{b}]$$

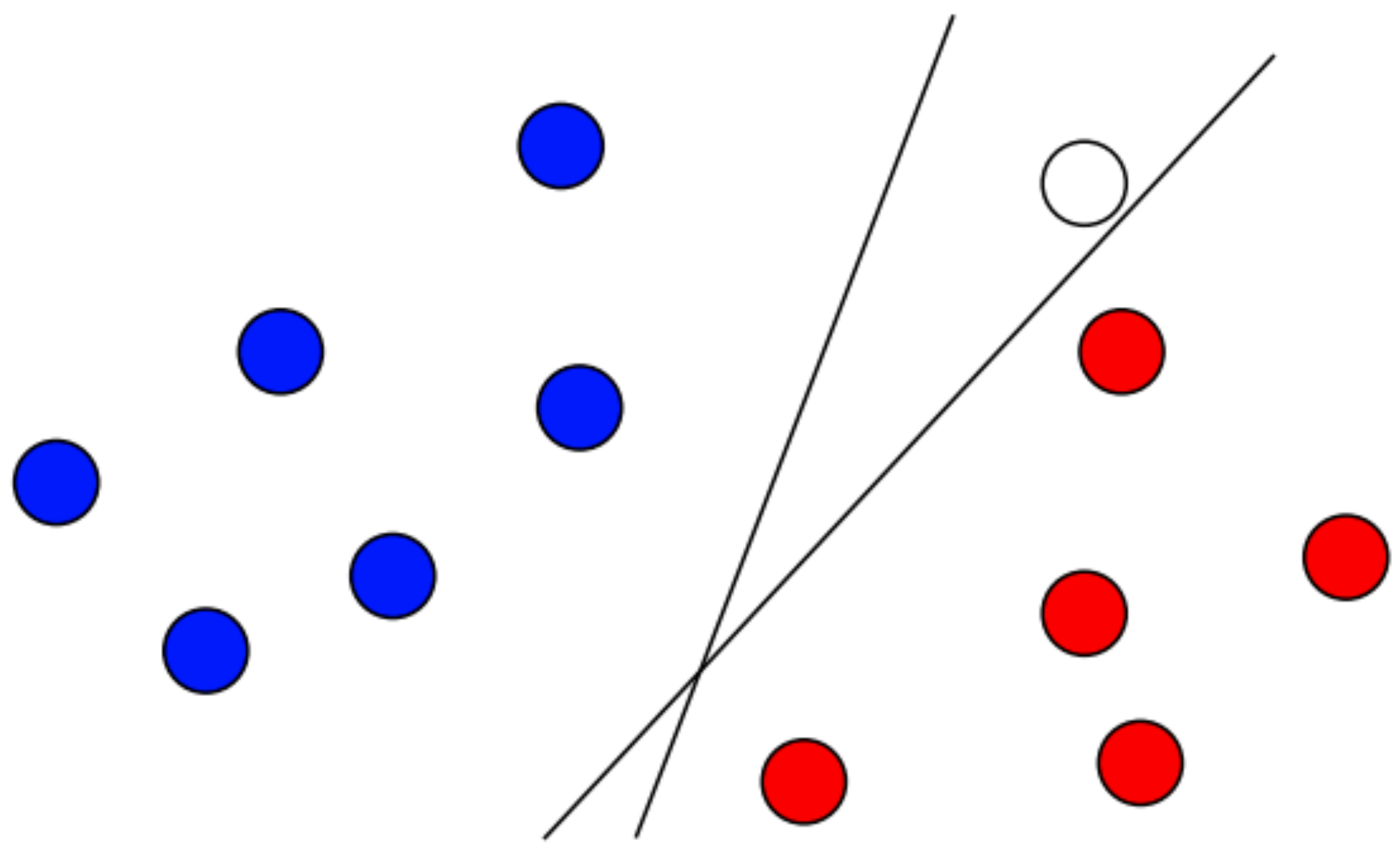
$$f_{\mathbf{w},b} = +1 \text{ if } \mathbf{w} \cdot \mathbf{x} + \text{b} > 0$$

$$f_{\mathbf{w},b} = -1 \text{ otherwise}$$

Note: $f_{\mathbf{w},b} = f_{\mathbf{w}',b'}$ if $\mathbf{w}' = c\mathbf{w}, b' = cb, \forall c > 0$

Support Vector Machines (SVM)

Which linear predictor is better?



* Illustration is taken from Julien Mairal presentation

Support Vector Machines (SVM)

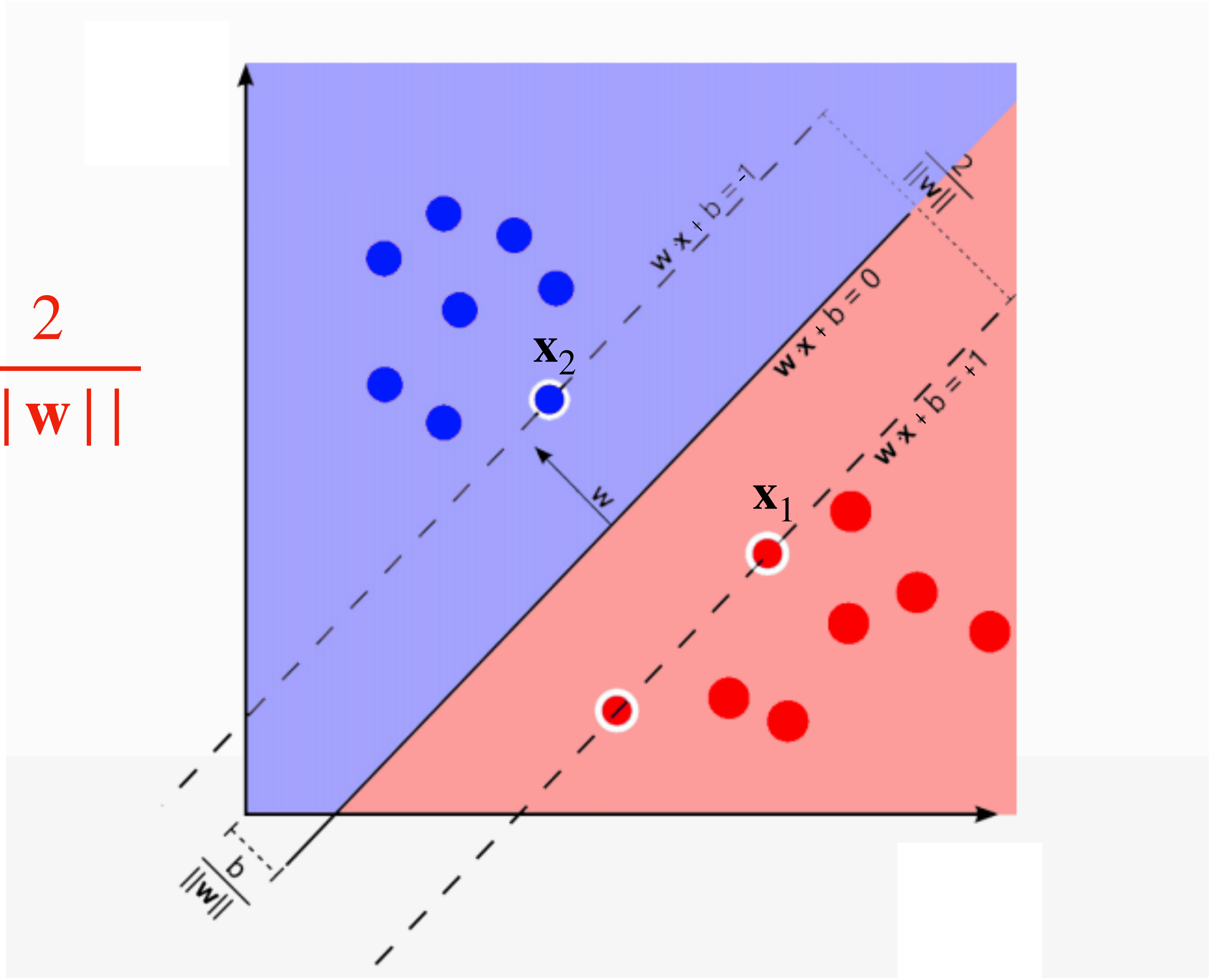
Margin

$$y(\mathbf{w} \cdot \mathbf{x} + b)$$

$$\mathbf{w} \cdot \mathbf{x}_1 + b = 1$$
$$\mathbf{w} \cdot \mathbf{x}_2 + b = -1$$

$$\mathbf{w} \cdot (\mathbf{x}_1 - \mathbf{x}_2) + b = 2$$

$$\frac{\mathbf{w}}{||\mathbf{w}||} \cdot (\mathbf{x}_1 - \mathbf{x}_2) + b = \frac{2}{||\mathbf{w}||}$$



Support Vector Machines (SVM)

All training points should be on the correct side of the dotted line

For $y_i = 1 \longrightarrow \mathbf{w} \cdot \mathbf{x}_i + b \geq 1$

For $y_i = -1 \longrightarrow \mathbf{w} \cdot \mathbf{x}_i + b \leq -1$

To summarize for both cases, we have

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \forall i \in \{1, \ldots, n\}$$

To find the optimal hyperplane

Find $(\mathbf{w}, b)$, which minimize $\frac{1}{2}||\mathbf{w}||^2$

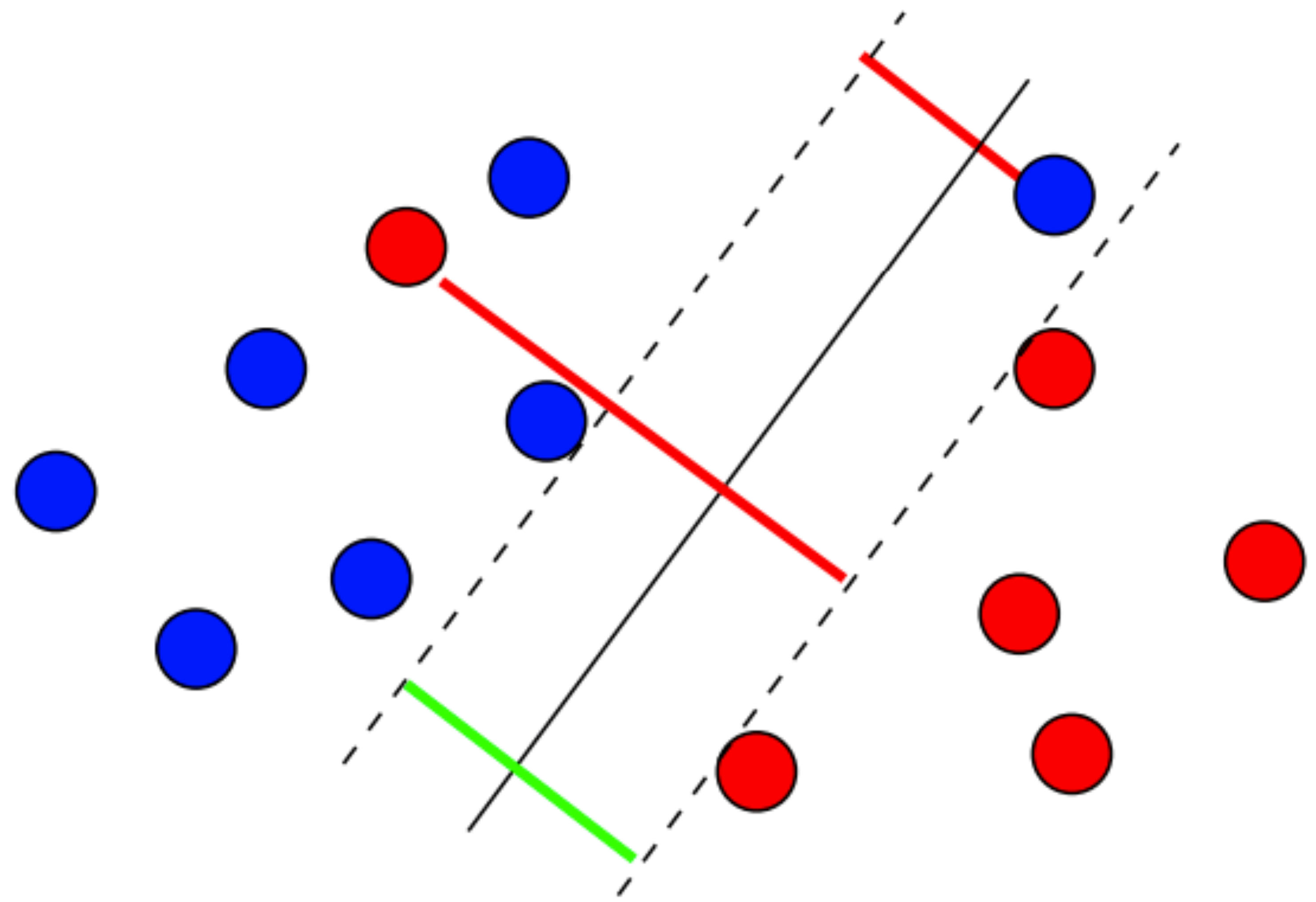
w.r.t. $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0, \forall i \in \{1, \ldots, n\}$

Data instances that satisfy $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$ called support vectors.

Support Vector Machines (SVM)

Soft margin

Data is not always linear separable -> Find a trade-off between large margin and errors



* Illustration is taken from Julien Mairal presentation

Support Vector Machines (SVM)

Margin

$$y(\mathbf{w} \cdot \mathbf{x} + b)$$

Error

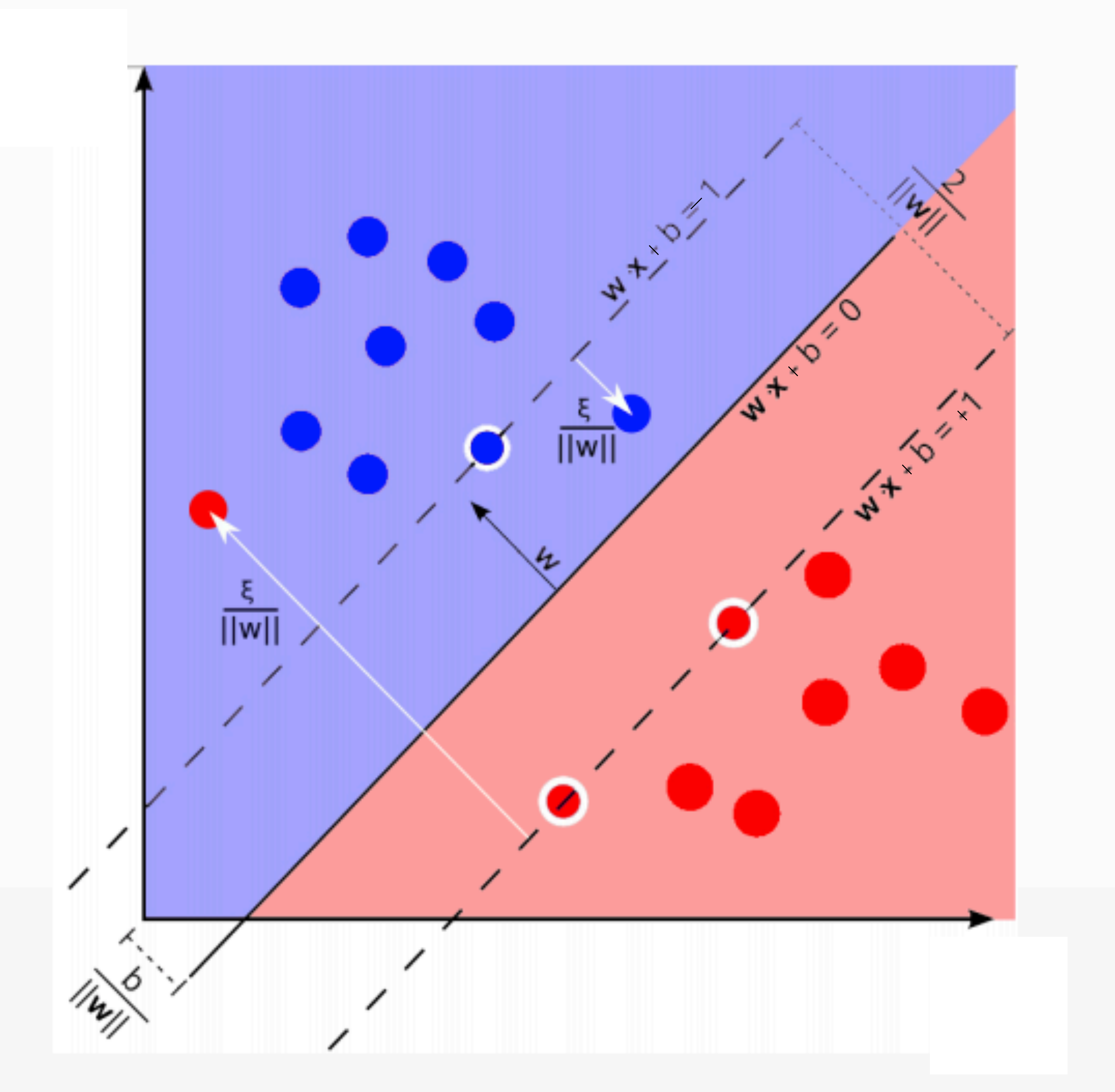
$$\xi = \max(0, 1 - y(\mathbf{w} \cdot \mathbf{x} + b))$$

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

w.r.t. $$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \forall i \in \{1, \ldots, n\}$$

Support Vector Machines (SVM)

- $\xi_i > 1$: an instance misclassified
- $0 < \xi_i < 1$: an instance is classified correctly, but lies inside the margin
- $\xi_i = 0$: an instance is classified correctly and lies outside of the margin.

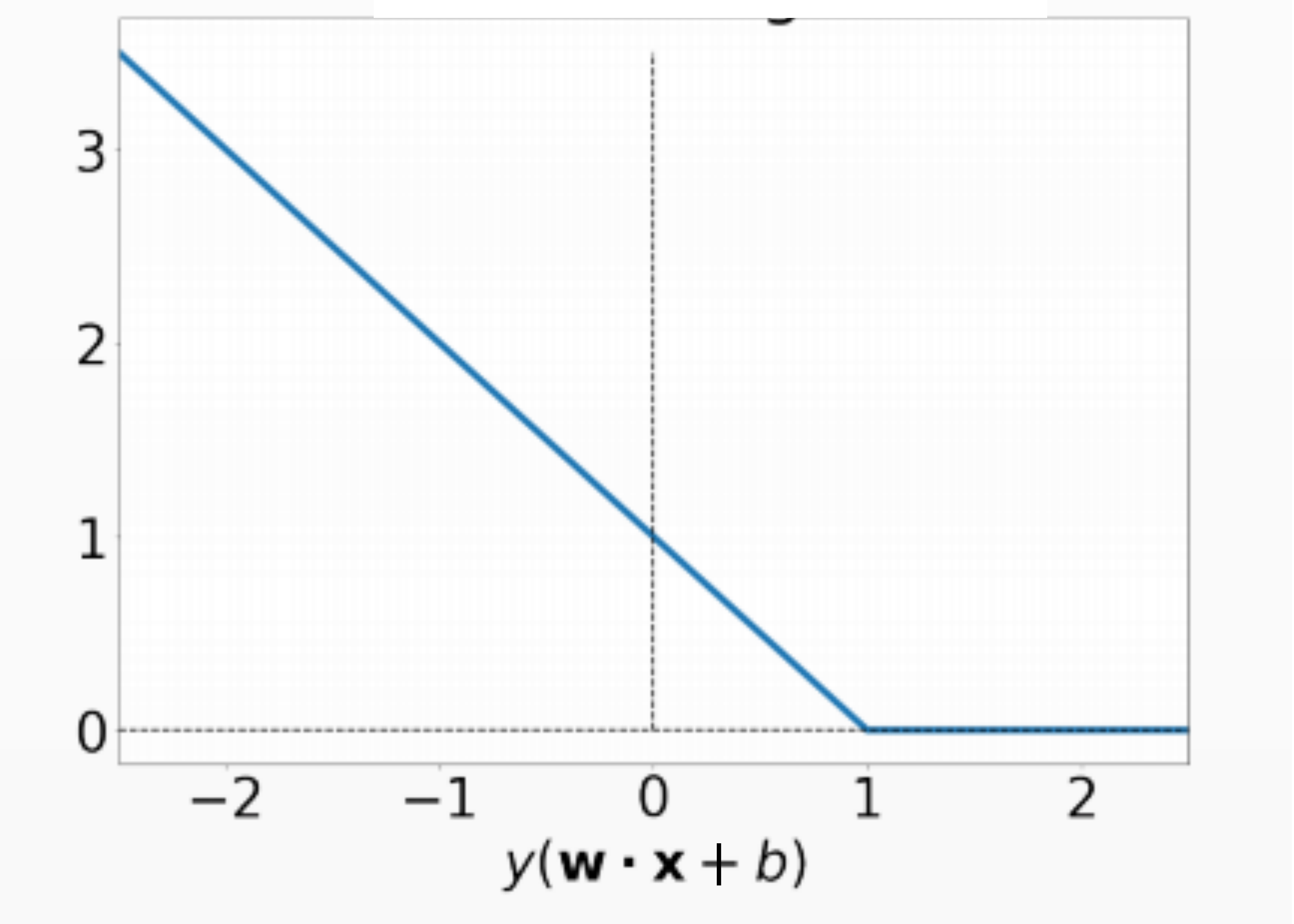


Support Vector Machines (SVM)

Minimize $$\frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n} l_{\text{hinge}}(f_{\mathbf{w},b}(\mathbf{x}_i, y_i))$$

where $$l_{\text{hinge}}(\hat{y}, y) = \max\{0, 1 - \hat{y} \times y\}$$ and $$\hat{y} = \mathbf{w} \cdot \mathbf{x} + b$$

Hinge loss



Loss functions

SVM

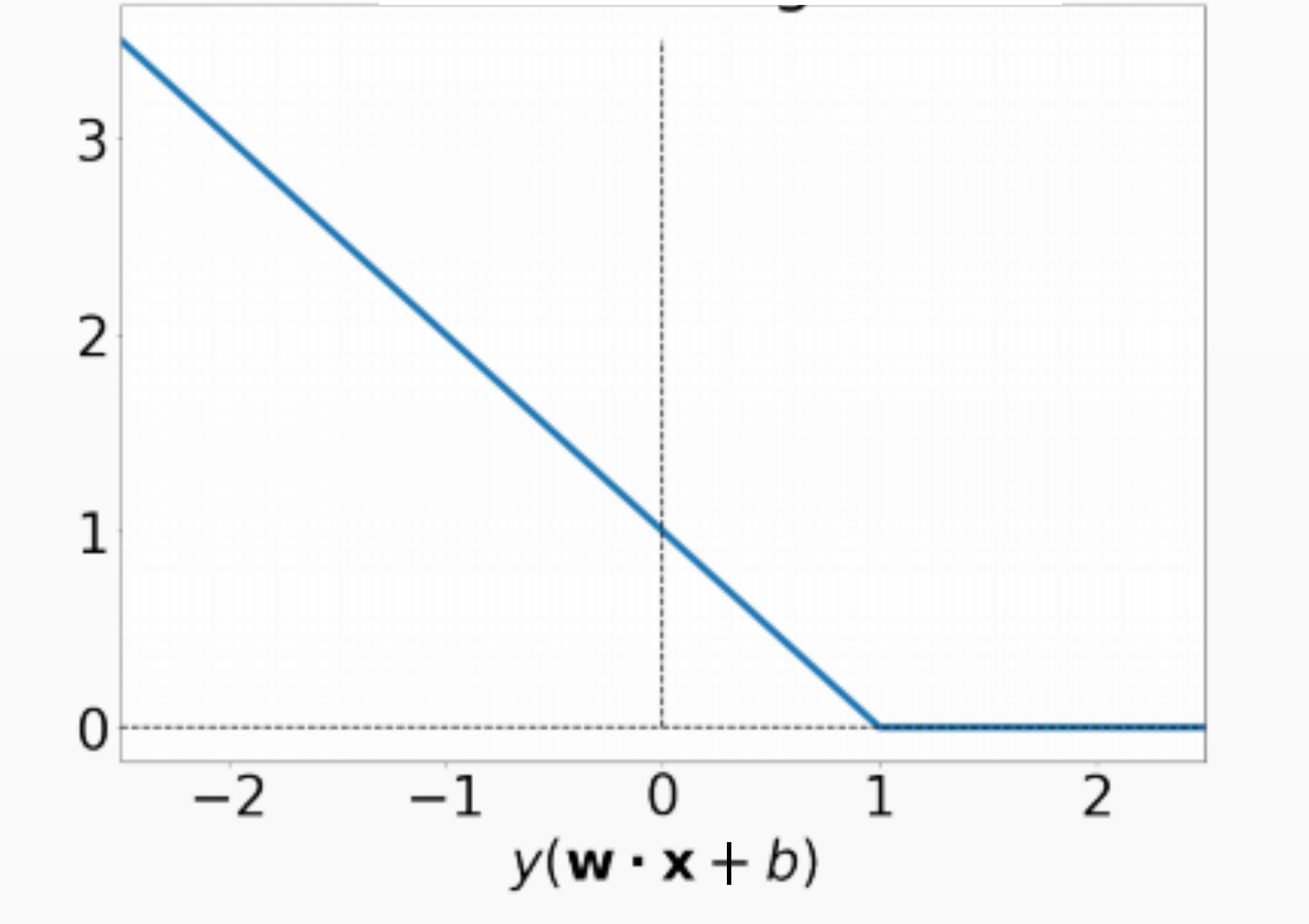
Minimize $$F(\mathbf{w}) = C\sum_{i=1}^{n} l_{\text{hinge}}(f_{\mathbf{w},b}(\mathbf{x}_i, y_i) + \frac{1}{2}||\mathbf{w}||^2$$

Empirical loss

Regularization

$l_{\text{hinge}}(\hat{y}, y) = \max\{0, 1 - \hat{y} \times y\}$ and $\hat{y} = \mathbf{w} \cdot \mathbf{x} + b$

Hinge loss



Loss functions

Logistic regression

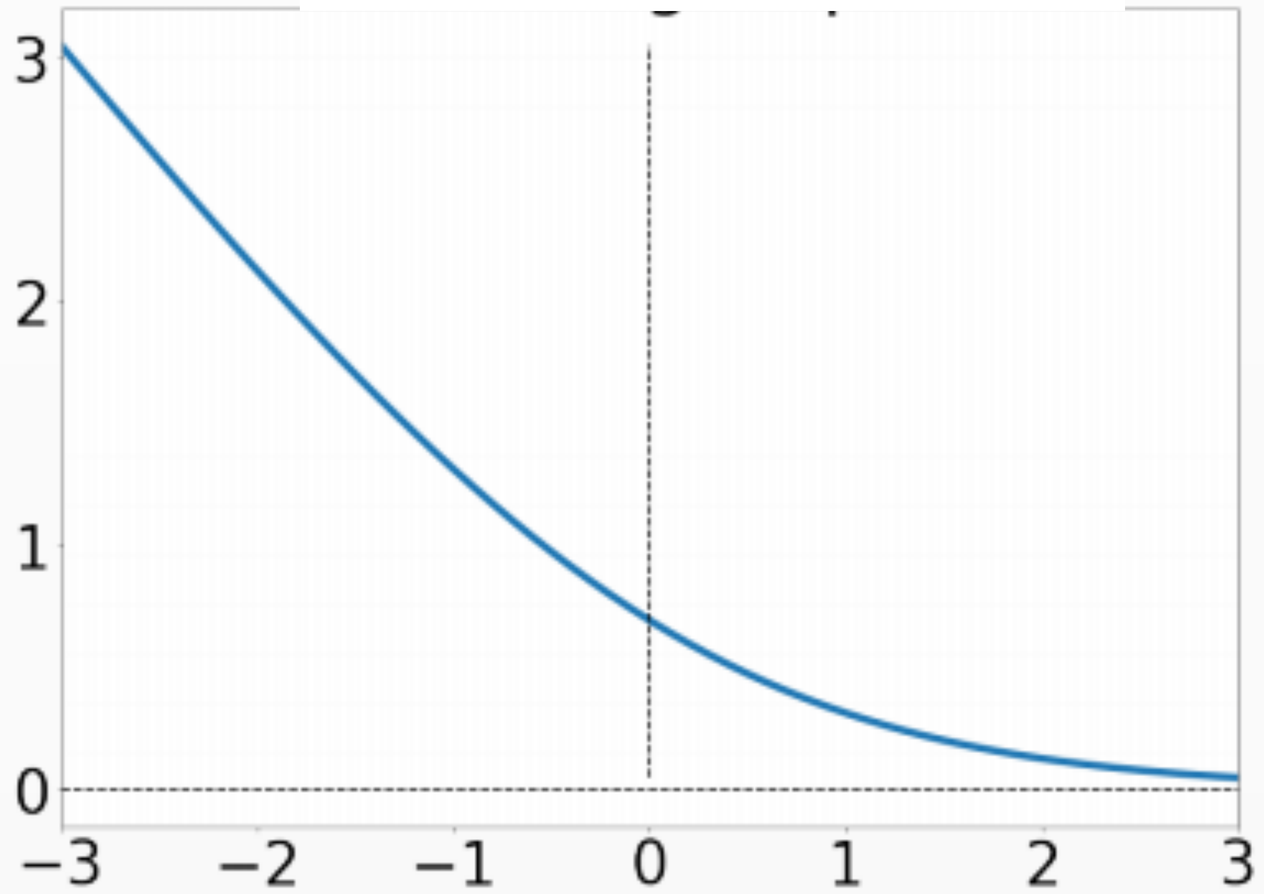
Minimize $$F(\mathbf{w}) = C\sum_{i=1}^{n} l_{\text{logist}}(f_{\mathbf{w},b}(\mathbf{x}_i), y_i) + \frac{1}{2}||\mathbf{w}||^2$$

Empirical loss

Regularization

$$l_{\text{logist}}(\hat{y}, y) = \ln(1 + e^{-\hat{y}\times y})$$ and $$\hat{y} = \mathbf{w}\cdot\mathbf{x} + b$$

Logistic loss



$y(\mathbf{w}\cdot\mathbf{x}+b)$