COS234:

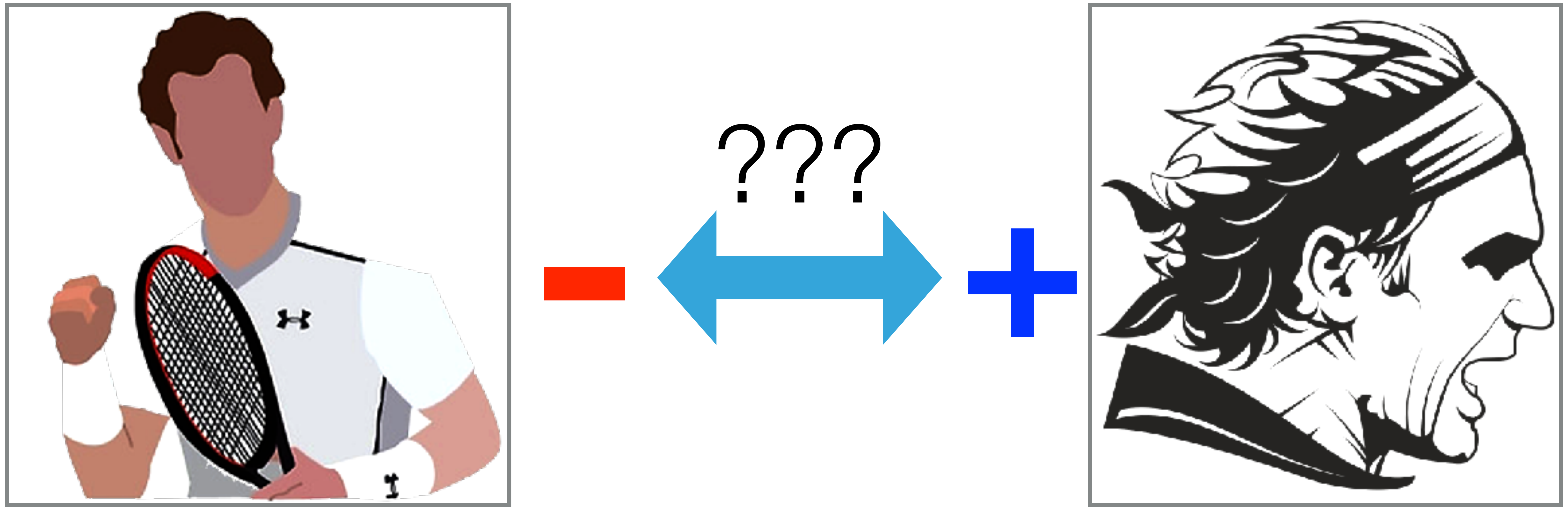
Introduction To Machine Learning

Prof. Yoram Singer

Princeton University

Topic: Linear Classification

Who Will Win Wimbledon?

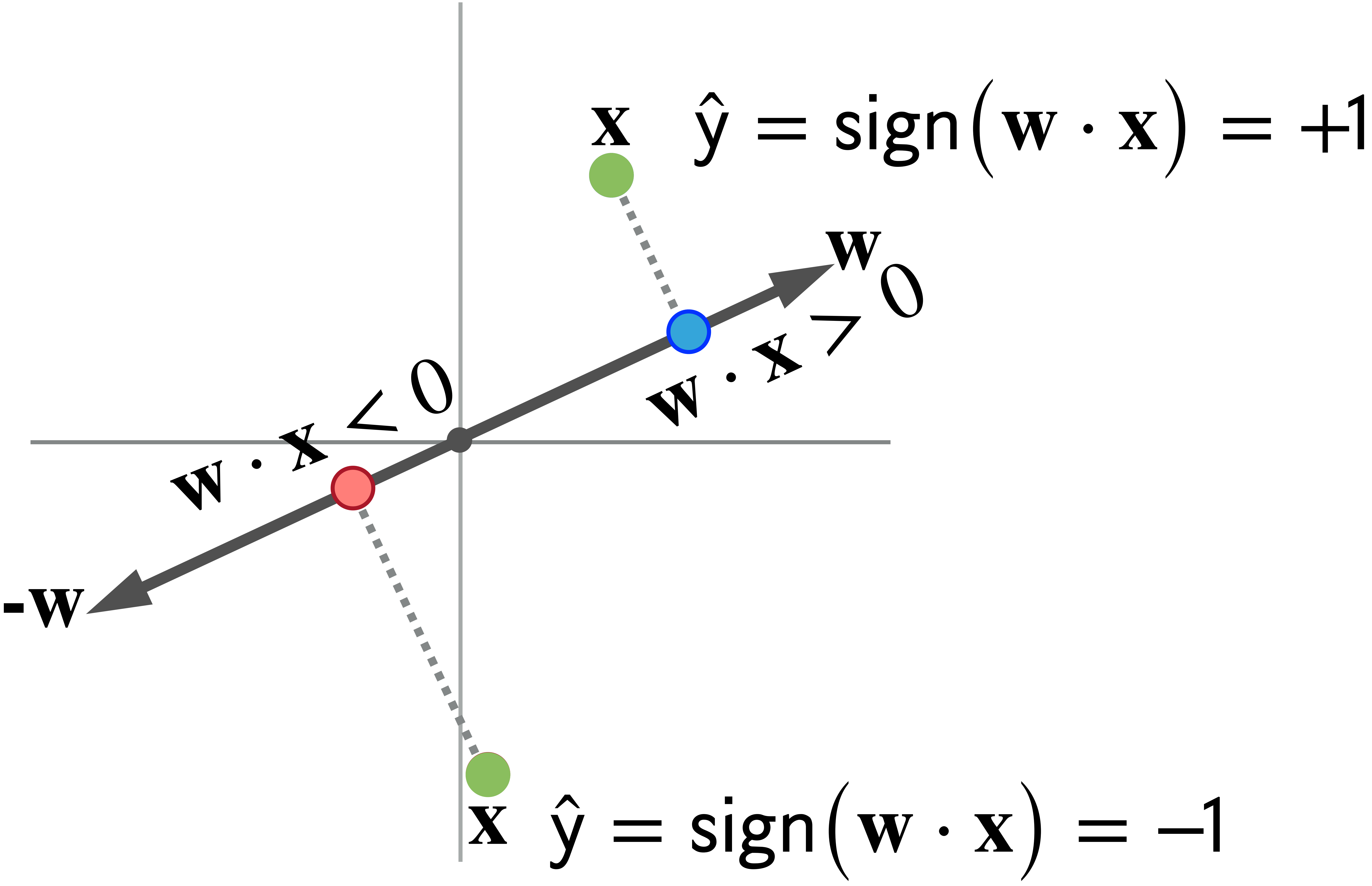


	H2H	Court	Injur.	Kids	Temp.	Rack.	
	⋮						⋮
$\mathbf{x_i}$	4 – 7	Grass	8 – 1	0 – 2	45F	8 –21	1
$\mathbf{x_{i+1}}$	1 – 1	Clay	1 - 0	0 - 0	81F	3 - 12	-1
	⋮						⋮

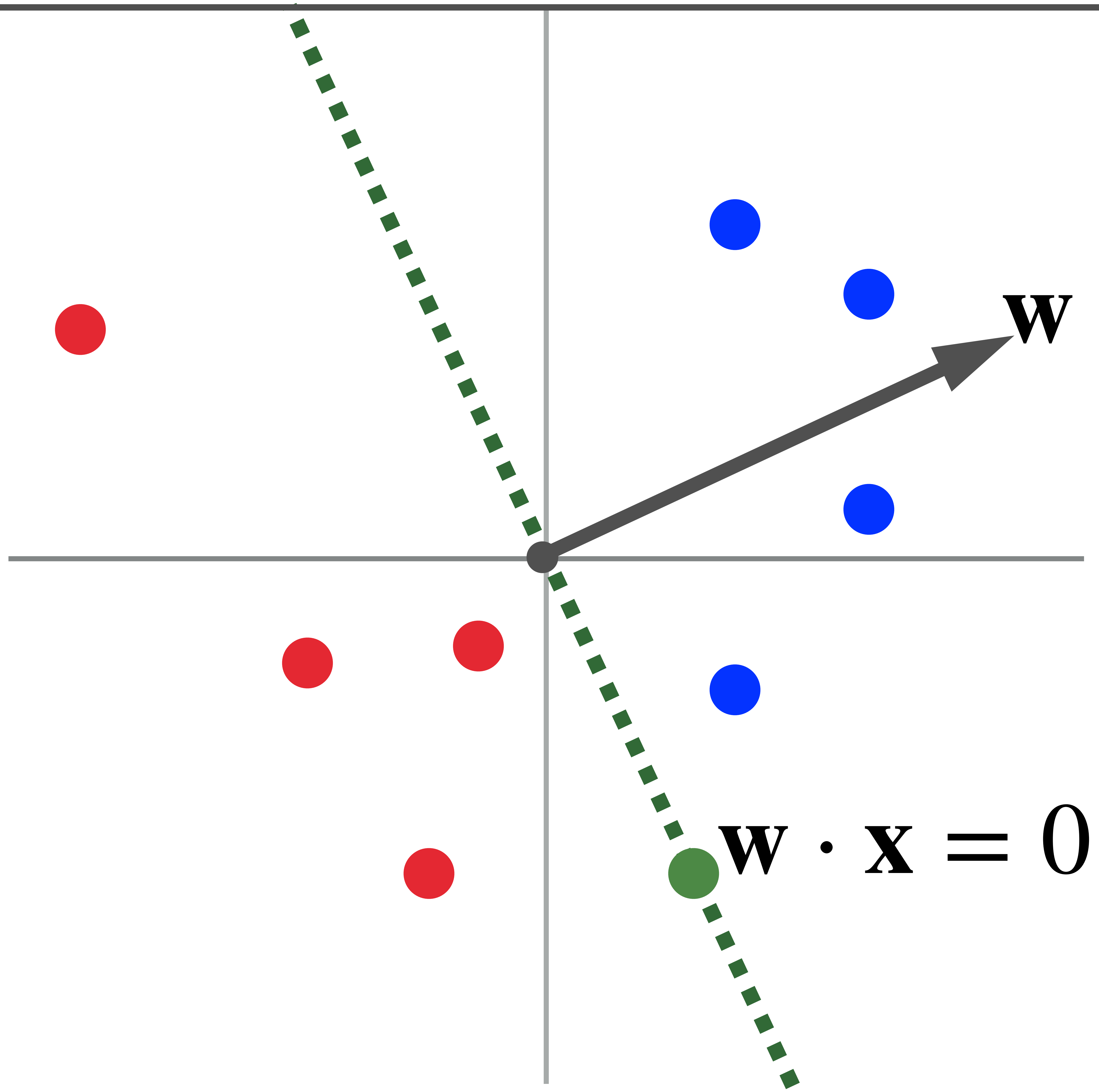
$\mathbf{w}$	1.7	-3	2.9	12	-5	0.1

Classification

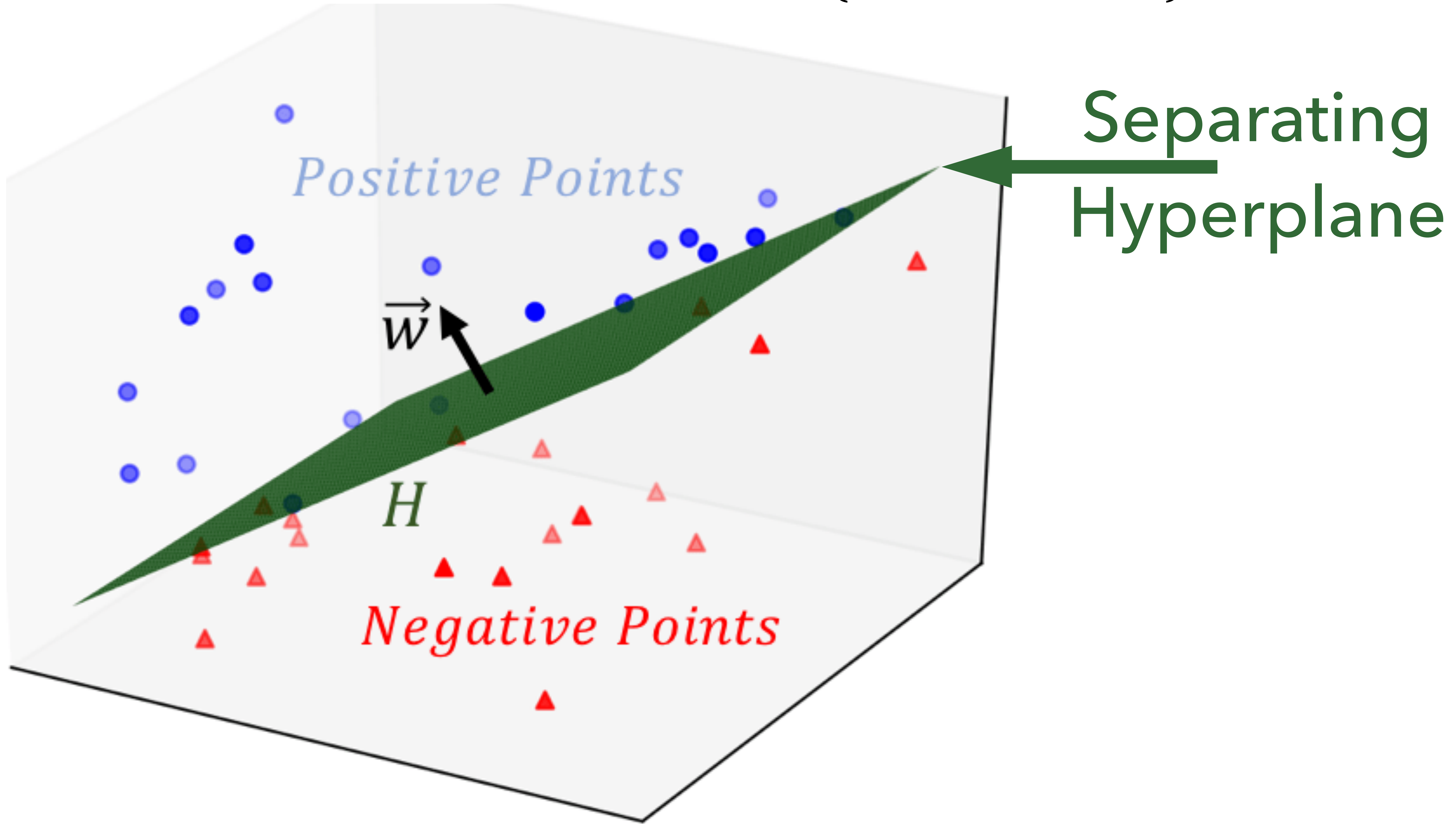
- If $\mathbf{w} \cdot \mathbf{x} > 0$ /* yes, True, positive */
  Predict **+1**
- If $\mathbf{w} \cdot \mathbf{x} < 0$ /* no, False, negative */
  Predict **-1**
- If $\mathbf{w} \cdot \mathbf{x} = 0$ /* toss a coin */
  Predict **-1** w.p **0.5** & **+1** w.p **0.5**
- Prediction is $\hat{y} = \mathbf{sign}(\mathbf{w} \cdot \mathbf{x})$


$\mathbf{x}$ $\hat{y} = \mathrm{sign}(\mathbf{w} \cdot \mathbf{x}) = +1$
$\mathbf{w}$
$\mathbf{w} \cdot \mathbf{x} > 0$
$\mathbf{w} \cdot \mathbf{x} < 0$
$\mathbf{-w}$
$\mathbf{x}$ $\hat{y} = \mathrm{sign}(\mathbf{w} \cdot \mathbf{x}) = -1$

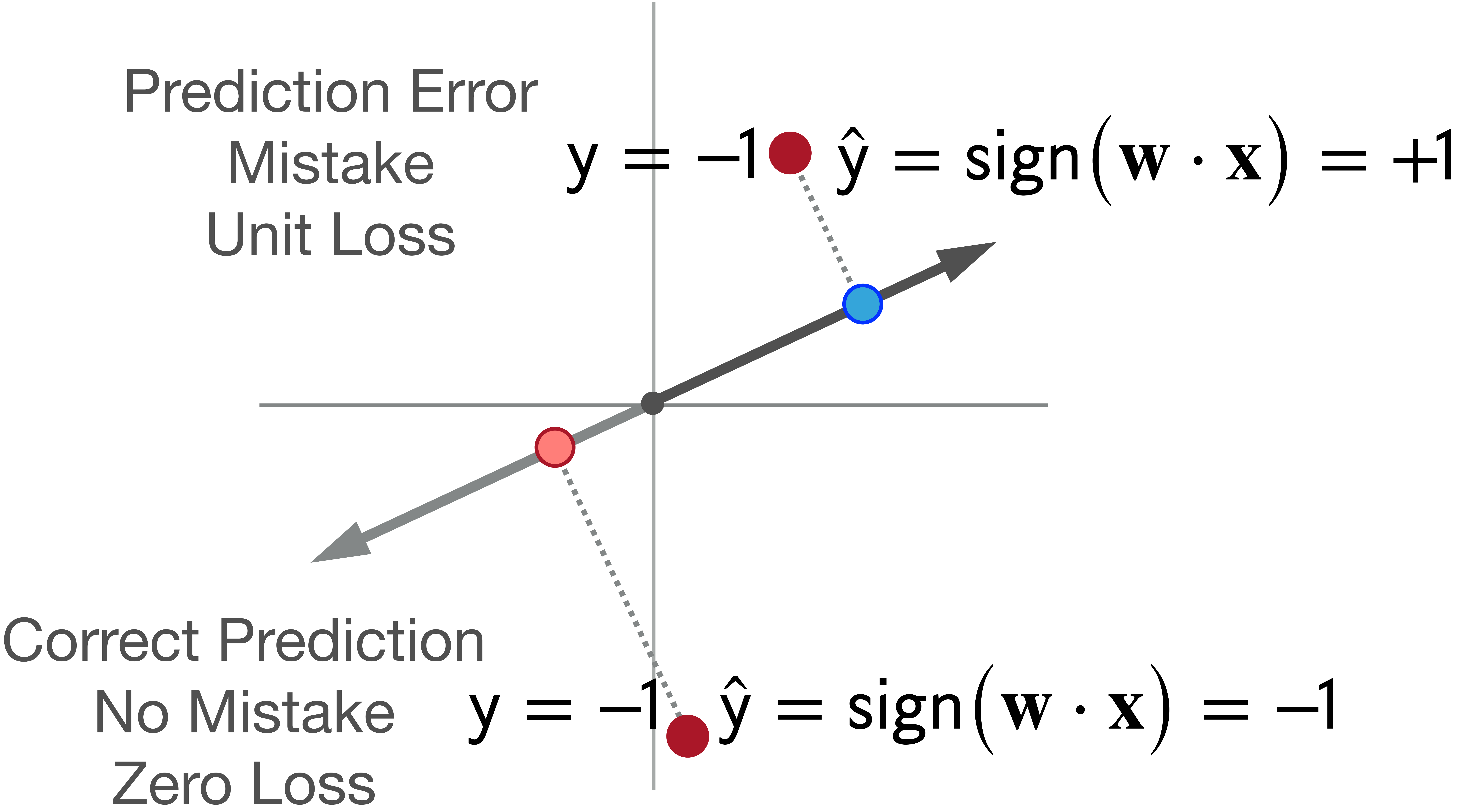
$\mathbf{w}$ is normal to hyperplane defined by $\{\mathbf{x} \mid \mathbf{w} \cdot \mathbf{x} = 0\}$



$\mathbf{w}$ is **normal** to **hyperplane** defined by $\{\mathbf{x} \,|\, \mathbf{w} \cdot \mathbf{x} = 0\}$



$\mathbf{w}$ is **d** **dimensional** $\Rightarrow$ **hyperplane** is **subspace** whose **dimension** is **d-1**

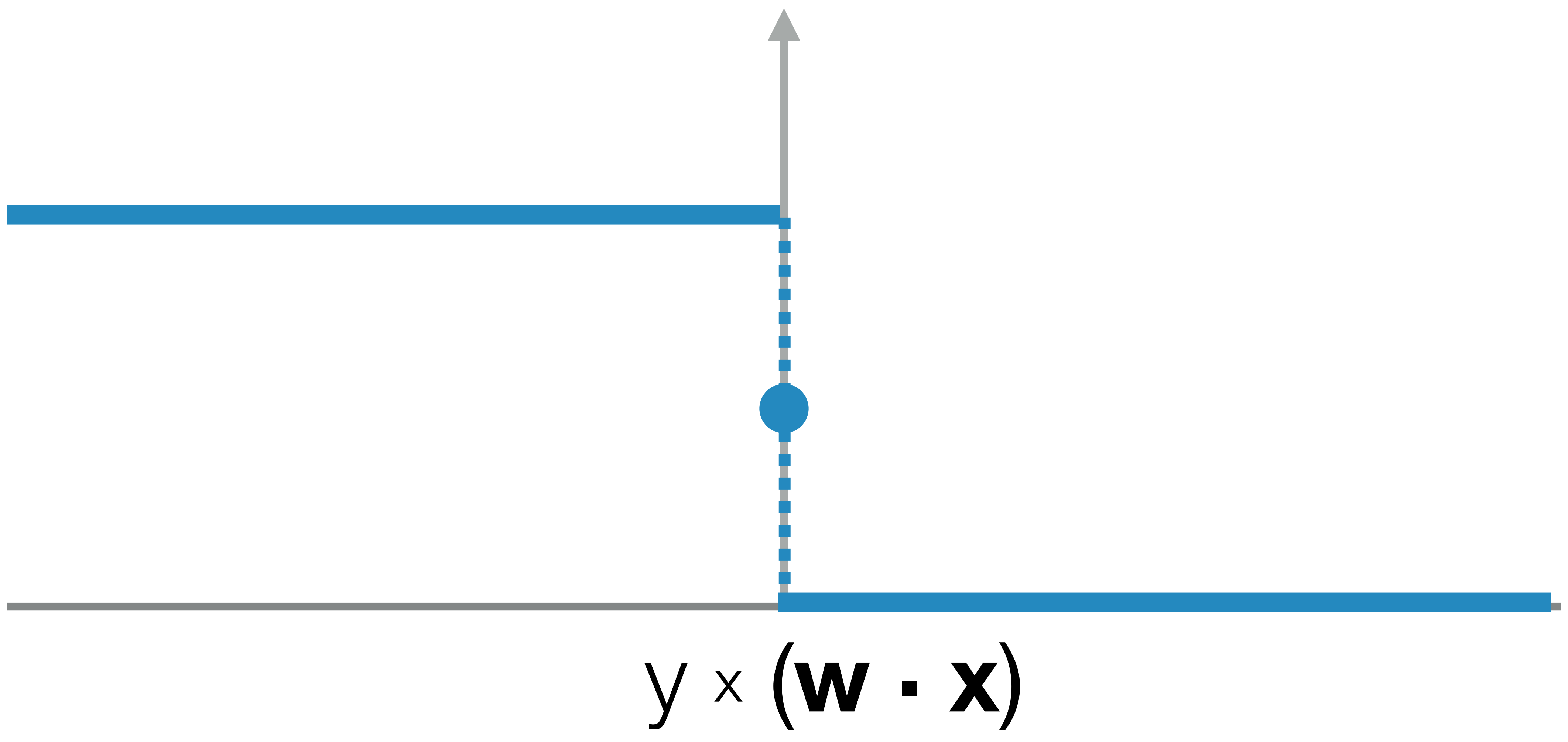

Prediction Error
Mistake
Unit Loss
$y = -1$ $\hat{y} = \mathrm{sign}(\mathbf{w} \cdot \mathbf{x}) = +1$
Correct Prediction
No Mistake
Zero Loss
$y = -1$ $\hat{y} = \mathrm{sign}(\mathbf{w} \cdot \mathbf{x}) = -1$

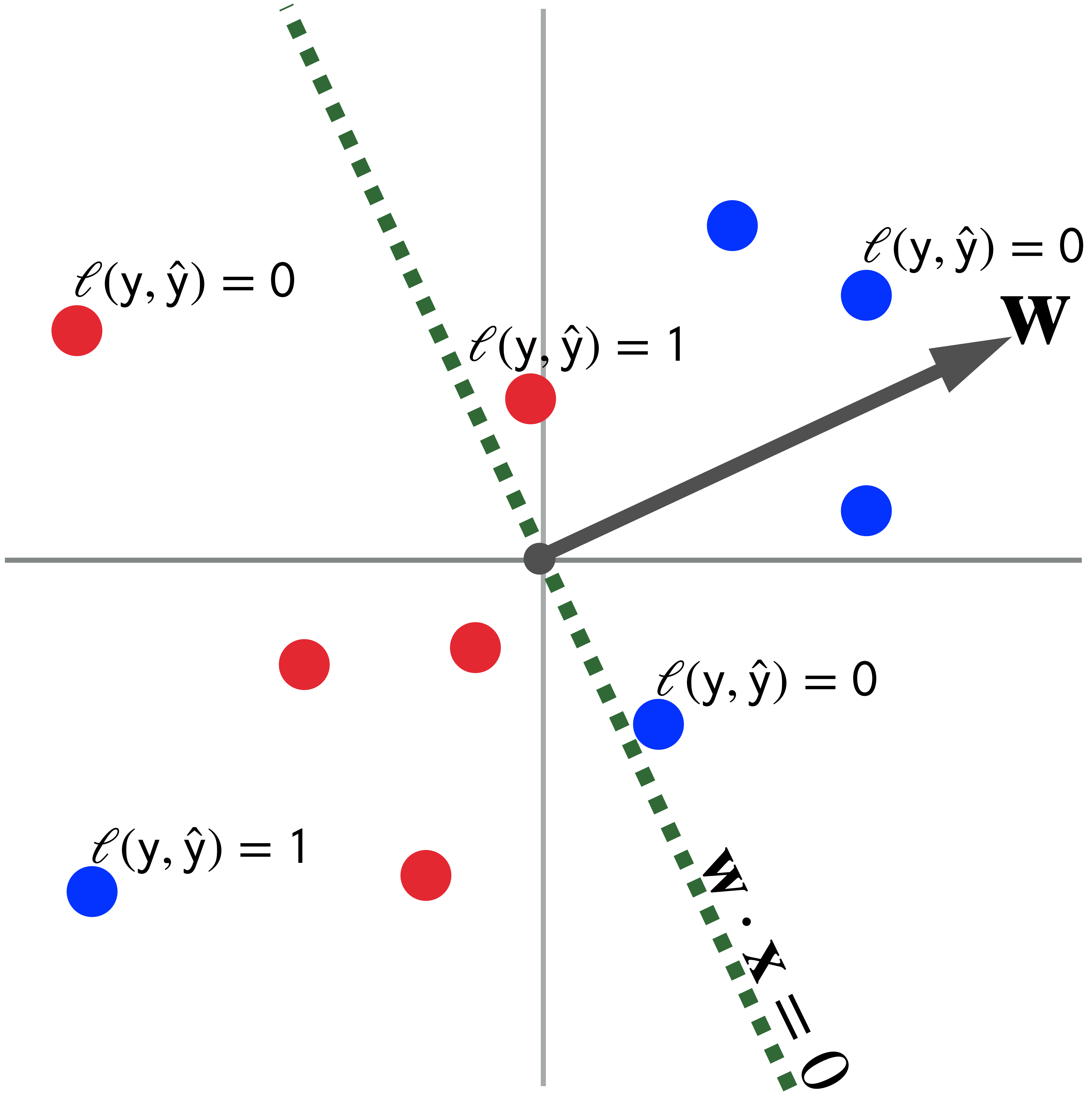
0-1 Loss

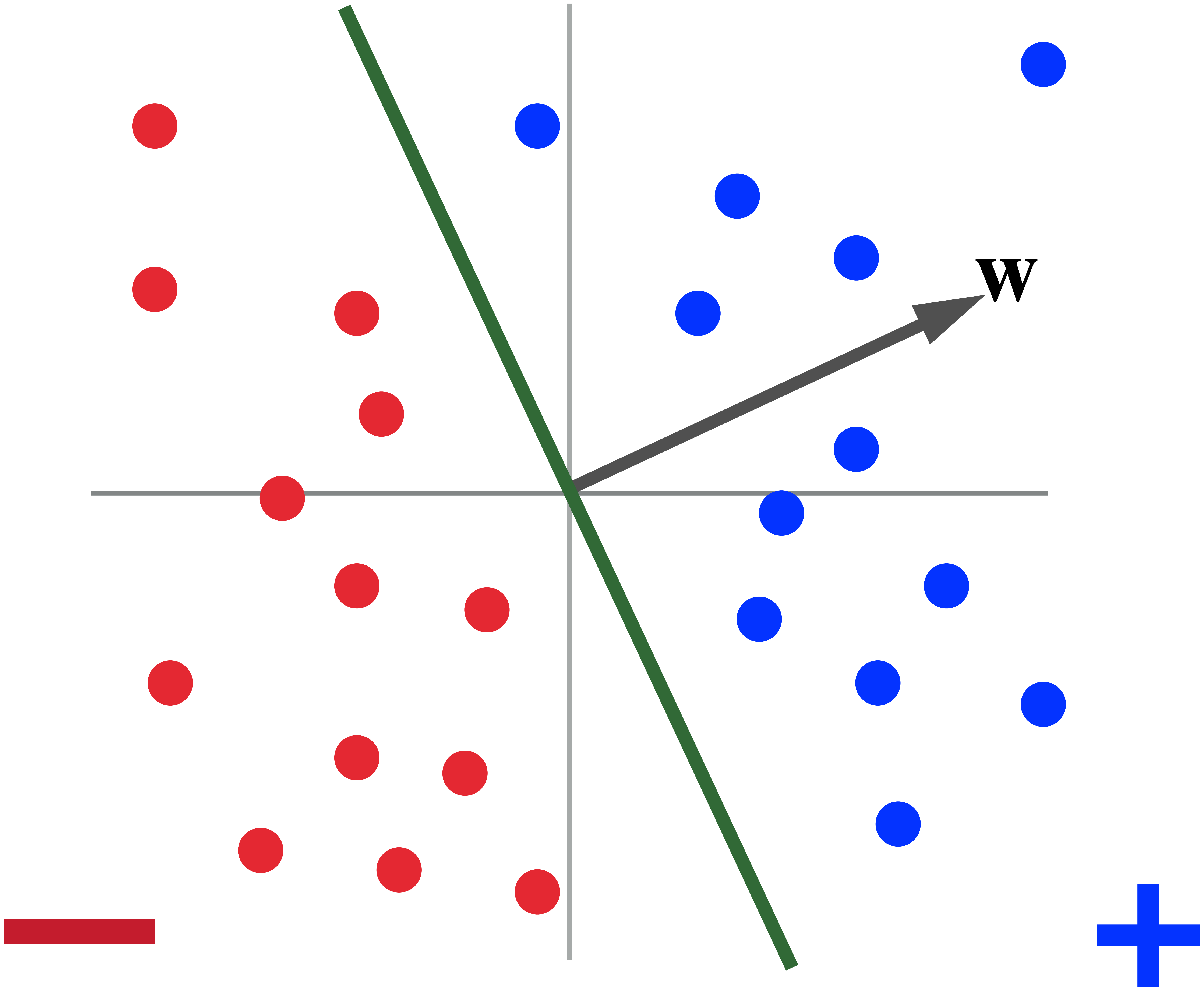
- For classification problems where $y, \hat{y} \in \{ -, + \}$ we could consider ℓ:

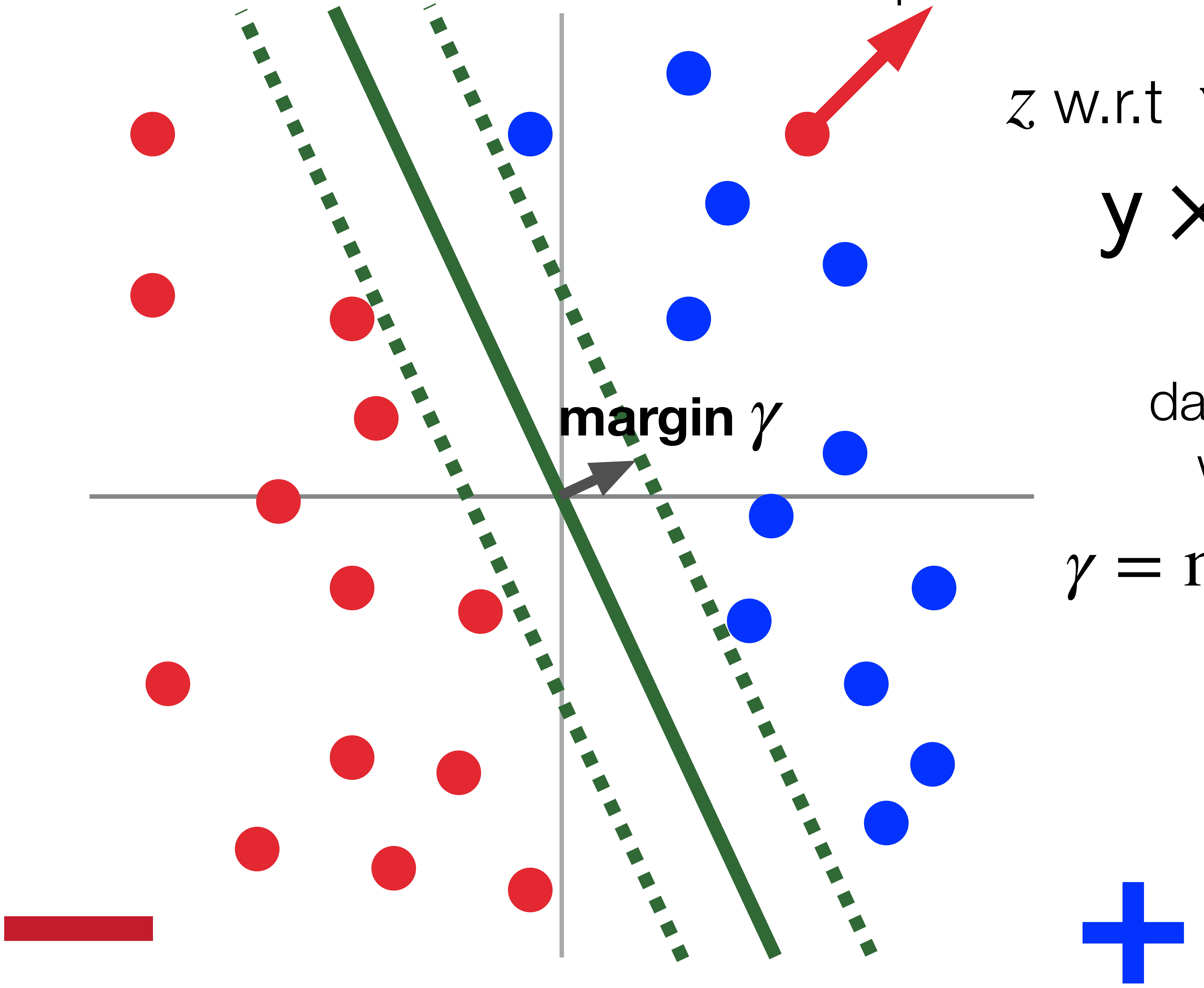
$$\ell(y, \hat{y}) \;=\; 1 - \frac{|\hat{y} + y|}{2} \;=\; \left[1 \text{ if } y \neq \hat{y} \text{ and } 0 \text{ o.w} \right] \;=\; \mathbf{1}\left[y \neq \hat{y}\right]$$

- Alas, this loss function is difficult to work with and <u>often</u> we need $\hat{y} \in \mathfrak{R}$ (why?)




$\ell(y, \hat{y}) = 0$
$\ell(y, \hat{y}) = 1$
$\ell(y, \hat{y}) = 0$
$\mathbf{w}$
$\ell(y, \hat{y}) = 0$
$\ell(y, \hat{y}) = 1$
$\mathbf{w} \cdot \mathbf{x} = 0$


w



not separable since $\gamma \leq 0$

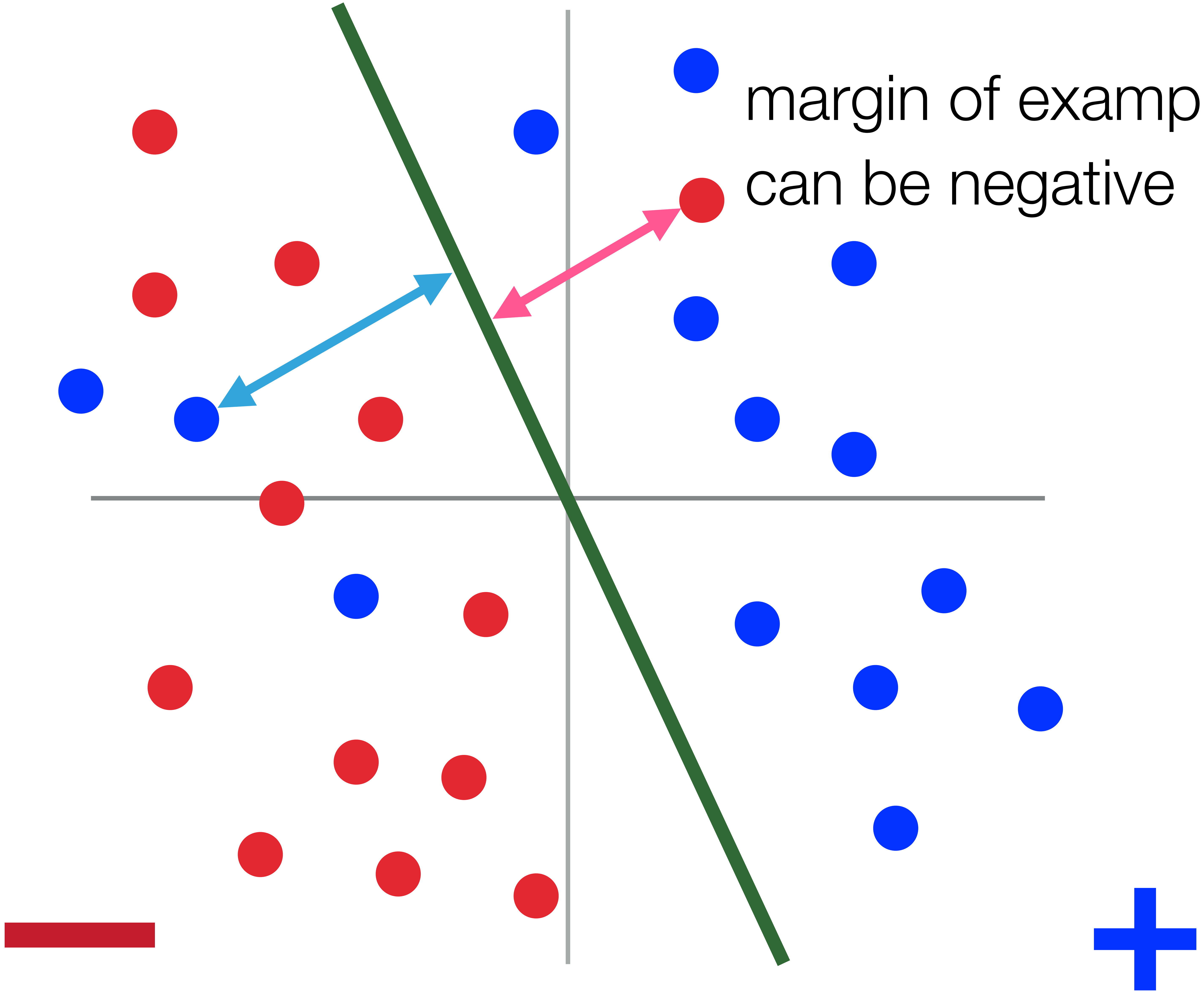
z w.r.t $\mathbf{w}$; $(\mathbf{x}, \mathrm{y})$ is:

$$\mathrm{y} \times (\mathbf{w} \cdot \mathbf{x})$$

data linearly-separable with margin $\gamma > 0$

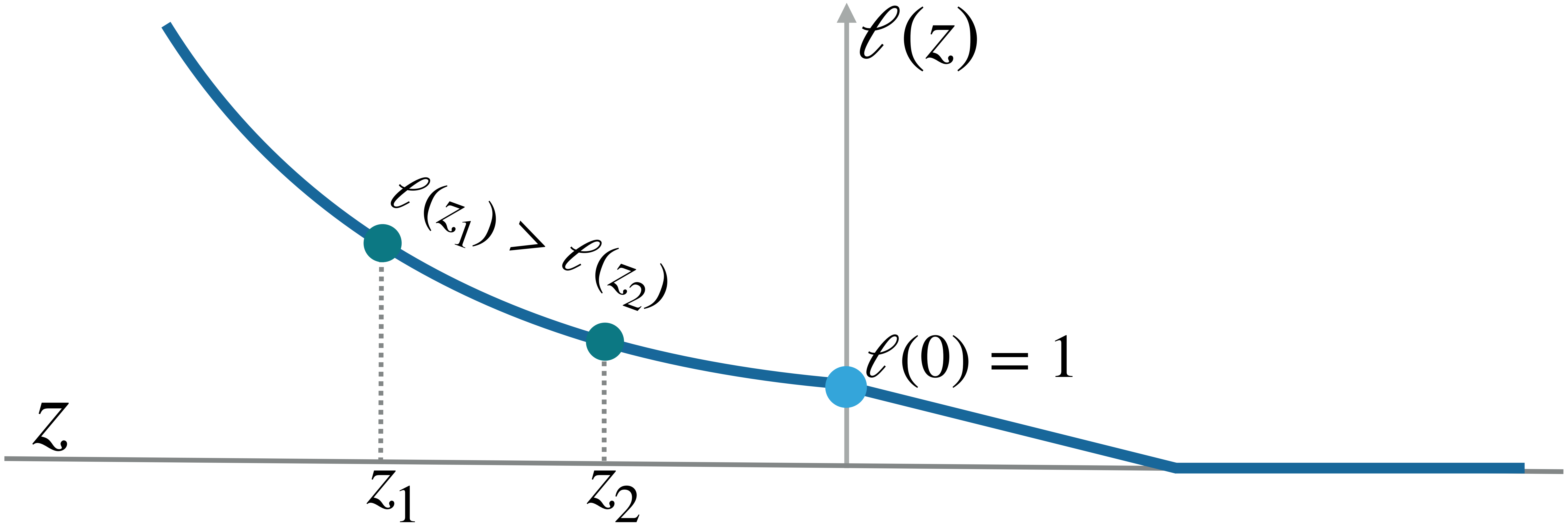
$$\gamma = \min_i z_i = \min_i y_i \, (\mathbf{w} \cdot \mathbf{x_i})$$

margin of example
can be negative

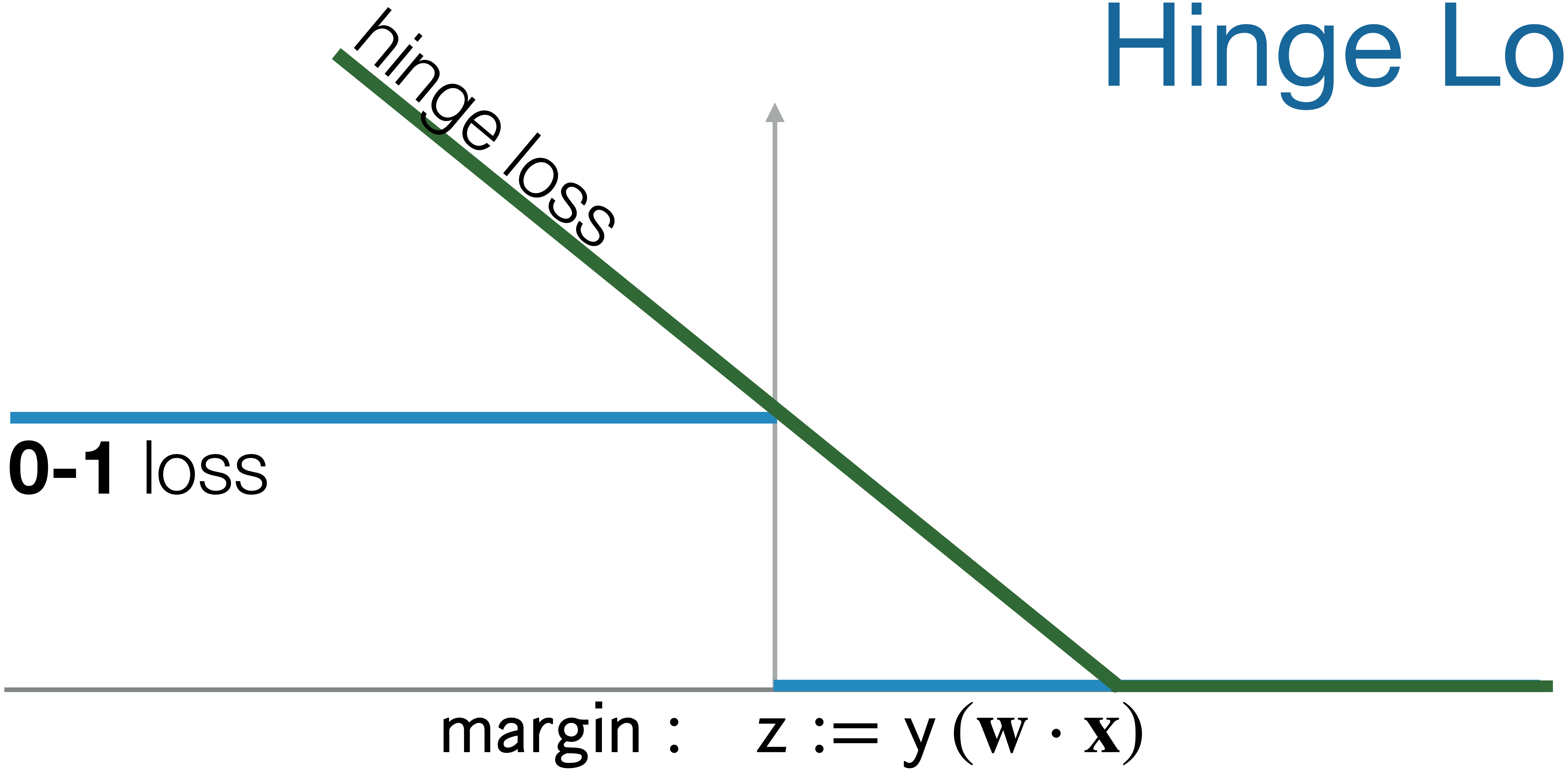


Margin-based Losses

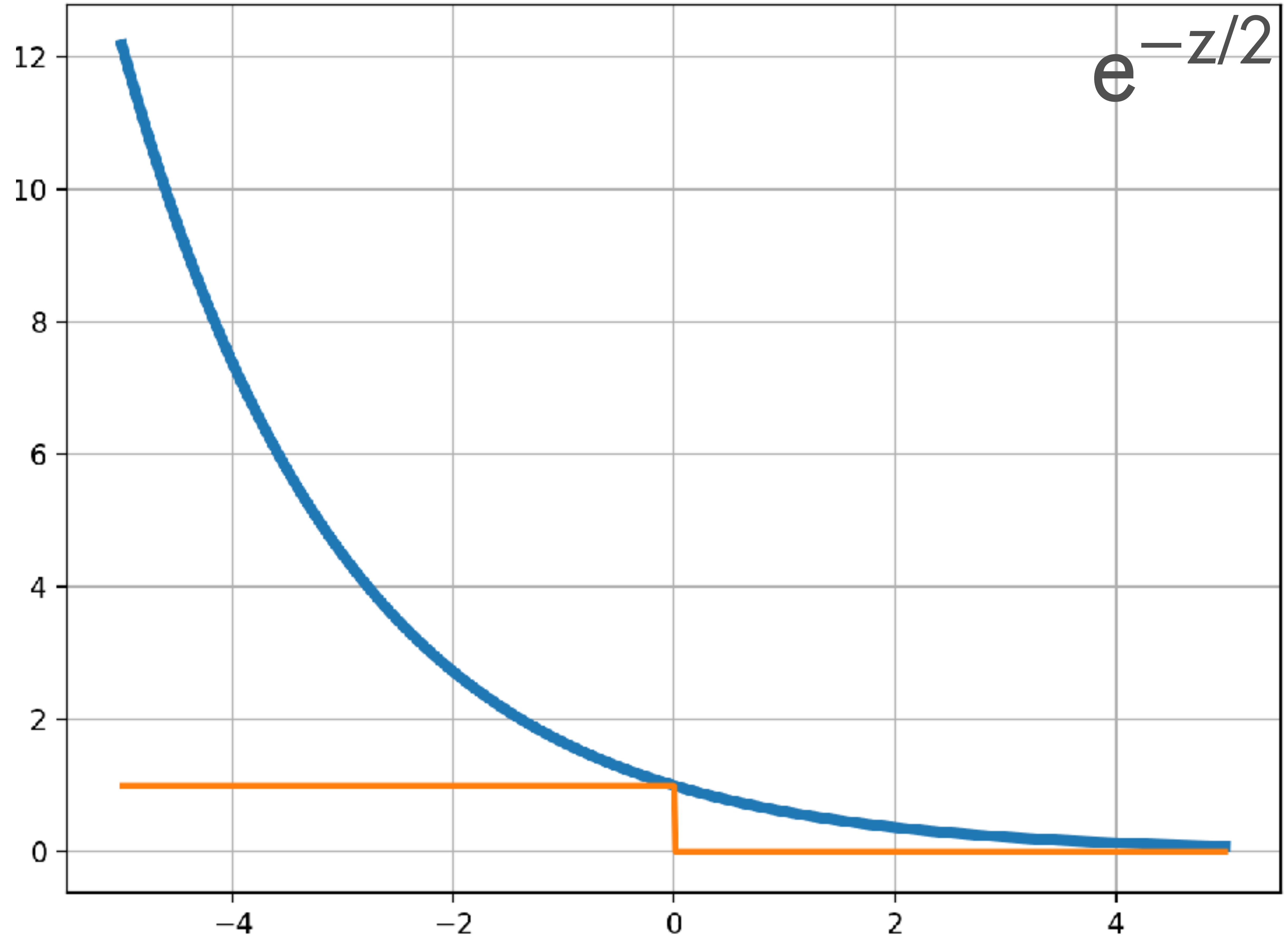
- $z := y\,(\mathbf{w}\cdot\mathbf{x}) = 0$ loss should be ≥ 1
- $z \to \infty$ loss should $\to 0$; $z \to -\infty$ loss should $\gg 0$ and often $\to \infty$
- If $z_1 < z_2$ we typically want $\ell(z_1) > \ell(z_2)$



Margin-Based Hinge Loss



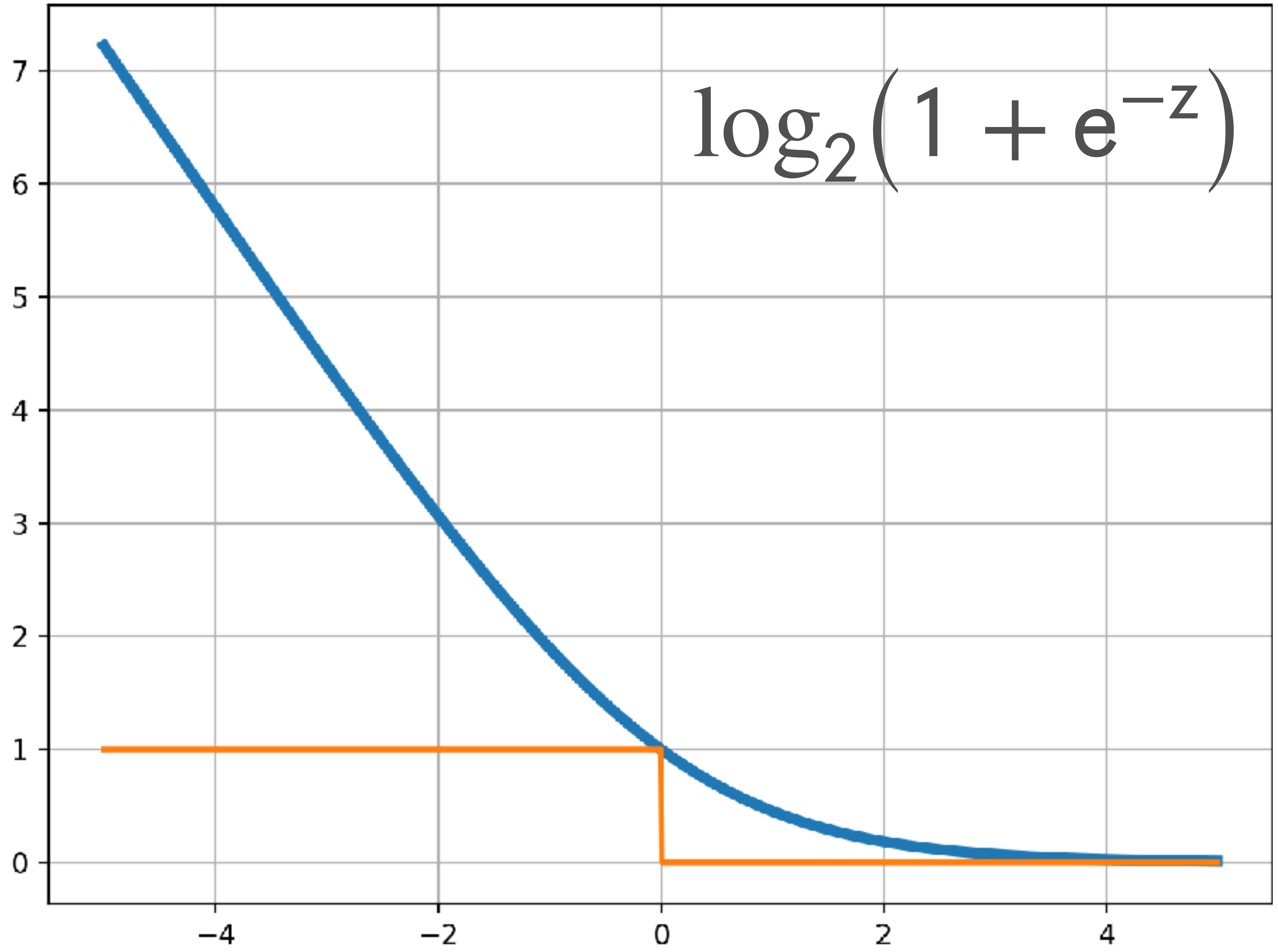
Exponential Loss



Logistic (Log) Loss

$$\log\left(1 + e^{-z}\right) \sim e^{-z} \quad \text{for} \quad z \gg 0$$

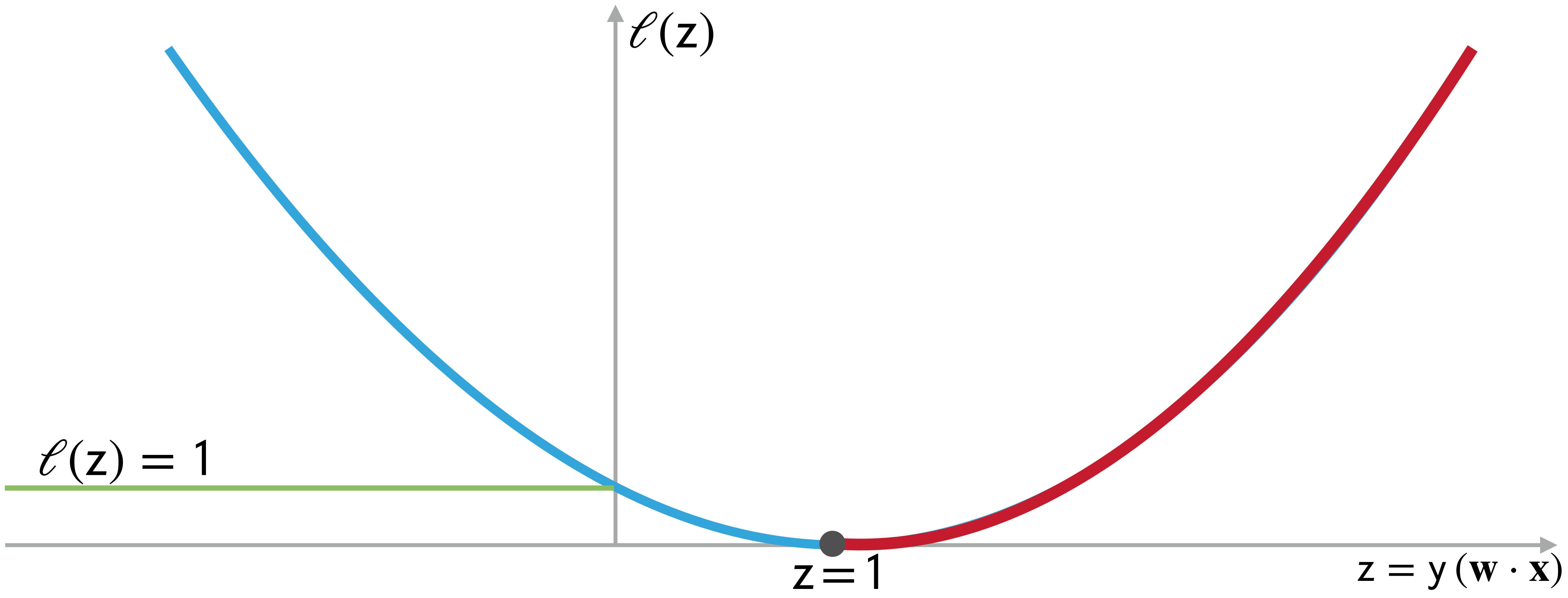
$$\log\left(1 + e^{-z}\right) \sim -z \quad \text{for} \quad z \ll 0$$



Squared Error?

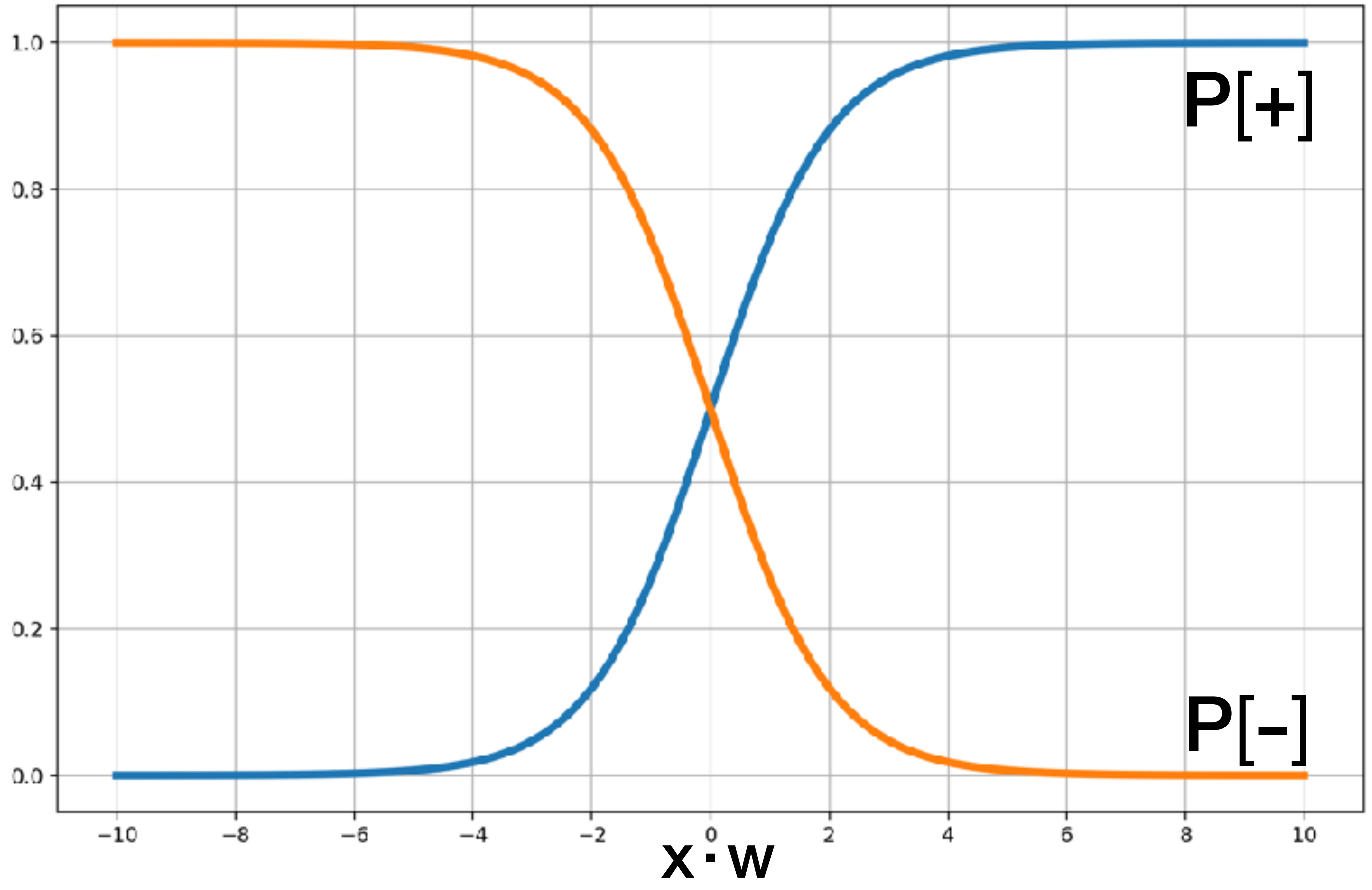
- Recall that $z = y \times (\mathbf{w}^T \mathbf{x})$
- Though y takes only two values {-1,+1} use as target
- Squared error:

$$\left(y - \mathbf{w}\cdot\mathbf{x}\right)^2 = y^2\left(y - \mathbf{w}\cdot\mathbf{x}\right)^2 = \left(y^2 - y\,\mathbf{w}\cdot\mathbf{x}\right)^2 = \left(1 - z\right)^2$$



Logistic Regression

- Given **x** "probability" of y to be +1: $\mathbf{P}\left[+1 \,|\, \mathbf{x}; \mathbf{w}\right] = \dfrac{1}{1 + e^{-\mathbf{w}\cdot\mathbf{x}}}$
- Probability of y to be -1: $\mathbf{P}\left[-1 \,|\, \mathbf{x}; \mathbf{w}\right] = 1 - \dfrac{1}{1 + e^{-\mathbf{w}\cdot\mathbf{x}}} = \dfrac{1}{1 + e^{\mathbf{w}\cdot\mathbf{x}}}$



Logistic Regression

- Given **x** "probability" of y to be +1: $\mathbf{P}\left[+\,|\,\mathbf{x};\mathbf{w}\right] = \dfrac{1}{1+e^{-\mathbf{w}\cdot\mathbf{x}}}$

- Probability of y to be -1: $\mathbf{P}\left[-\,|\,\mathbf{x};\mathbf{w}\right] = 1 - \dfrac{1}{1+e^{-\mathbf{w}\cdot\mathbf{x}}} = \dfrac{1}{1+e^{\mathbf{w}\cdot\mathbf{x}}}$

- Combine two cases: $\mathbf{P}\left[\mathrm{y}\,|\,\mathbf{x};\mathbf{w}\right] = \dfrac{1}{1+e^{-\mathrm{y}\,\mathbf{w}\cdot\mathbf{x}}} = \dfrac{1}{1+e^{-\mathrm{z}}}$

  Predict +1 w.p. P[+|**x**] and -1 w.p. P[-|**x**]

- Define loss to be negative of log-probability (log-likelihood):

$$-\log\left(\mathbf{P}\left[\mathrm{y}\,|\,\mathbf{x};\mathbf{w}\right]\right) = \log\left(1+\mathrm{e}^{-\mathrm{z}}\right)$$