(proof contied:) On the other hand,

$$\|w^{(k)}\|^2 \overset{\text{update}}{=} \|w^{(k-1)} + y_i x^{(i)}\|^2 \quad \left(x^{(i)} \text{ is misclassified} \right)$$

$$\overset{\text{def of inner prod}}{=} \left(w^{(k-1)} + y_i x^{(i)}, \; w^{(k-1)} + y_i x^{(i)} \right)$$

$$= \left(w^{(k-1)}, w^{(k-1)} \right) + \left(y_i x^{(i)}, y_i x^{(i)} \right)$$

$$+ 2 \left(w^{(k-1)}, y_i x^{(i)} \right)$$

$$\overset{\text{def of inner prod}}{=} \|w^{(k-1)}\|^2 + \underbrace{\|x^{(i)}\|^2}_{\leq R^2}$$

$$+ \underbrace{2 y_i \left(x^{(i)} \cdot w^{(k-1)} \right)}_{< 0 \text{ since } x^{(i)} \text{ is misclassified by } w^{(k-1)}}$$

$$\leq \|w^{(k-1)}\|^2 + R^2$$

Similarly: $\|w^{(k-1)}\|^2 \leq \|w^{(k-2)}\|^2 + R^2$ add

$$\vdots$$

$$\|w^{(1)}\|^2 \leq \underbrace{\|w^{(0)}\|^2}_{=0} + R^2 = R^2$$

Add the inequalities to get

$$\|w^{(k)}\|^2 \leq kR^2.$$

(upper bound of $\|w^{(k)}\|$)

Combine the lower and upper bounds:

$$k^2\gamma^2 \leq \|w^{(k)}\|^2 \leq kR^2$$

$$\Rightarrow \quad k^2\gamma^2 \leq kR^2$$

$$\Rightarrow \quad k \leq \frac{R^2}{\gamma^2}. \qquad \blacksquare$$

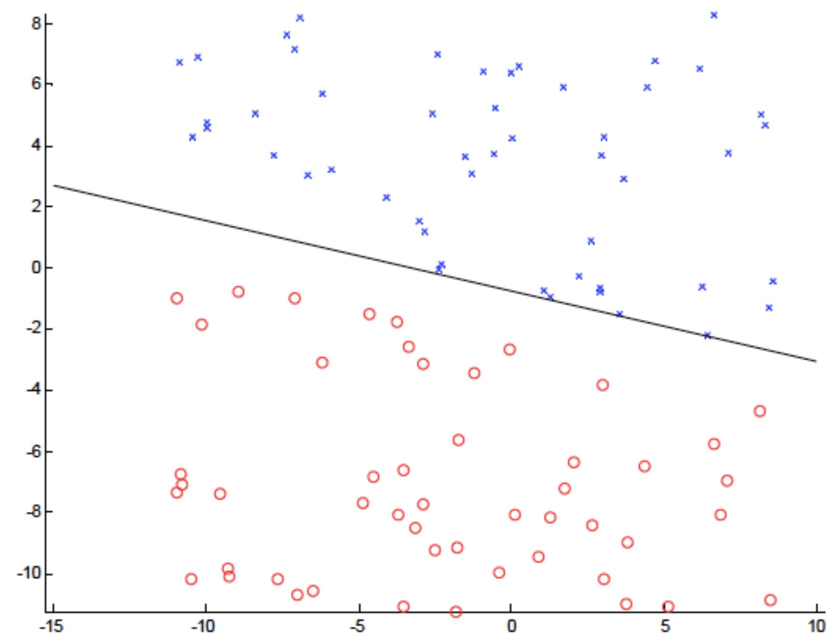
Summary of this section:

- What is a perception? Def.
- How to implement? Algorithm.

Reference: "Foundations of Data Science"
by A. Blum, J. Hopcroft, R. Kannan
Section 5.2.

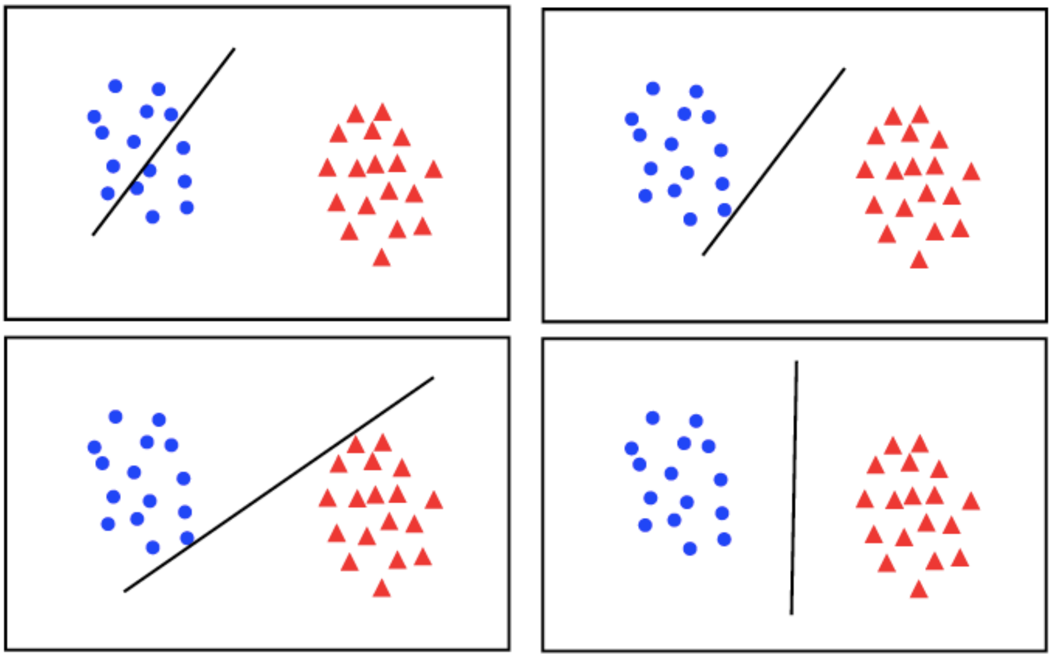
4.2 Support Vector Machine

4.2.1 Motivation.



- if the data is linearly separable, then the algorithm will converge
- convergence can be slow …
- separating line close to training data

What is the best w?



• maximum margin solution: most stable under perturbations of the inputs

Observation: classification using hyperplanes with the maximum margin is most stable with respect to perturbations of the inputs.

Def of Support Vector Machine (SVM): a SVM is a linear binary classifier based on the margin maximization.

Setting: Let $(x^{(1)}, y_1), \cdots, (x^{(n)}, y_n)$ be sample pts with labels $y_i = \pm 1$.

Denote $x^{(i)}$ by "$\times$" if $y_i = 1$,

by "$\circ$" if $y_i = -1$