パーセプトロンの収束

- LE CONVERGENCE DE LA PERCEPTRON -

パーセプトロン

入力ベクトル

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$\boldsymbol{x}$

$x \in \mathbb{R}^d$

出力ラベル

+1

or

-1

y

パーセプトロンの中身

入力ベクトル

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$\boldsymbol{x}$

$\boldsymbol{x} \in \mathbb{R}^d$

重みベクトル

$$\begin{cases} +1: & \boldsymbol{x} \cdot \boldsymbol{w} \geq 0 \\ -1: & \boldsymbol{x} \cdot \boldsymbol{w} < 0 \end{cases}$$

$\boldsymbol{w} \in \mathbb{R}^d$

出力ラベル

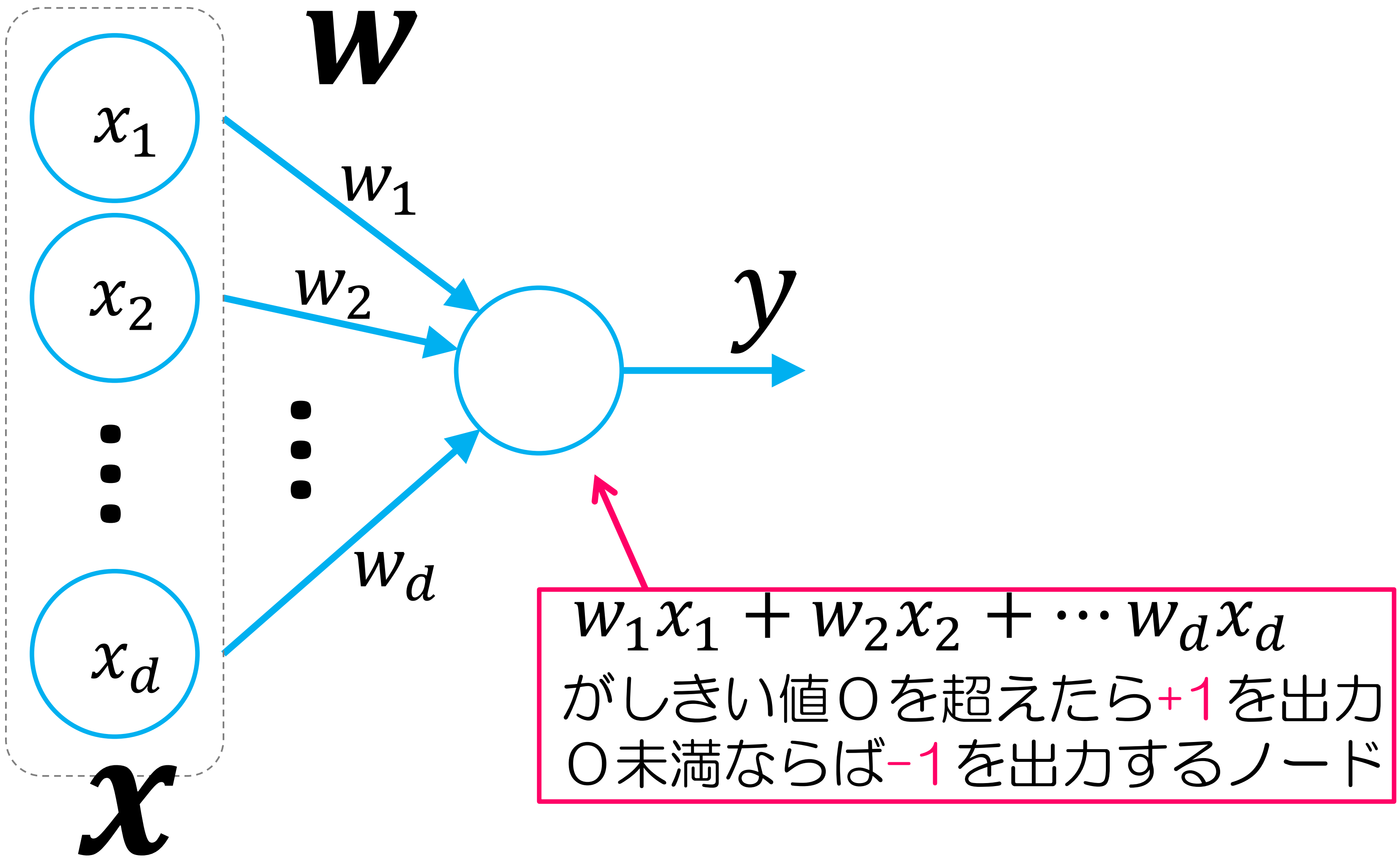
+1

or

-1

y

よくある図



$\boldsymbol{w}$
x_1
x_2
x_d
$\boldsymbol{x}$
w_1
w_2
w_d
y
$w_1 x_1 + w_2 x_2 + \cdots w_d x_d$
がしきい値０を超えたら+1を出力
０未満ならば−1を出力するノード

パーセプトロンの学習

訓練事例集合の中の訓練事例を前から順番に１つずつ取り出し，入力ベクトル$\mathbf{x}_i$をパーセプトロンに入力.
y_iと同じラベルが出力できれば何もしない.
y_iと同じラベルが出力できなければ重みベクトルを即時更新する

- **訓練事例：**
  - $(\mathbf{x}, y)$
    - $\mathbf{x}$: 入力ベクトル
    - y: 出力ラベル
- **訓練事例集合：**
  - $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots (\mathbf{x}_n, y_n)\}$

パーセプトロンの学習

訓練事例集合の中の訓練事例を前から順番に1つずつ取り出し，
入力ベクトル$\mathbf{x}_i$をパーセプトロンに入力．
y_iと同じラベルが出力できれば何もしない．
y_iと同じラベルが出力できなければ重みベクトルを即時更新する

オンライン学習

- **訓練事例：**
  - $(\mathbf{x}, y)$
    - $\mathbf{x}$: 入力ベクトル
    - y: 出力ラベル
- **訓練事例集合：**
  - $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots (\mathbf{x}_n, y_n)\}$

パーセプトロン学習アルゴリズム

t：重みの変更回数
i：今見ている事例の添え字
η：学習率（η>0）

- （初期化）
  - $\mathbf{w}^{(0)} = \mathbf{0}$
- （学習）
  - 重みベクトル$\mathbf{w}^{(t)}$で訓練事例集合中の要素を全て正しく分類できるまで以下を繰返し実行
    - for i=1 to n　//訓練事例を1～nまで１つずつ選択
      - if　$y_i(\mathbf{w}^{(t)} \cdot \mathbf{x}_i) \leqq 0$:　//分類に失敗したとき //括弧の中身と$y_i$は正負が逆
        - $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} + \eta y_i \mathbf{x}_i$　//重みの更新
      - else:　pass

パーセプトロンの更新式

$$if\ y_i(\boldsymbol{w}^{(t)} \cdot \boldsymbol{x}_i) \leq 0:$$

$$\boldsymbol{w}^{(t+1)} = \boldsymbol{w}^{(t)} + \eta y_i \boldsymbol{x}_i$$

- この式で本当にうまく学習が行えているのだろうか？

t：重みの変更回数
i：今見ている事例の添え字
η：学習率（η>0）

- **確認：**
  - 更新後の$\mathbf{w}^{(t+1)}$をもう一度 if文の条件式に入れてみると…

$$y_i(\boldsymbol{w}^{(t+1)} \cdot \boldsymbol{x}_i) = y_i(\boldsymbol{w}^{(t)} + \eta y_i \boldsymbol{x}_i) \cdot \boldsymbol{x}_i)$$

$$= y_i(\boldsymbol{w}^{(t)} \cdot \boldsymbol{x}_i + \eta y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}_i))$$

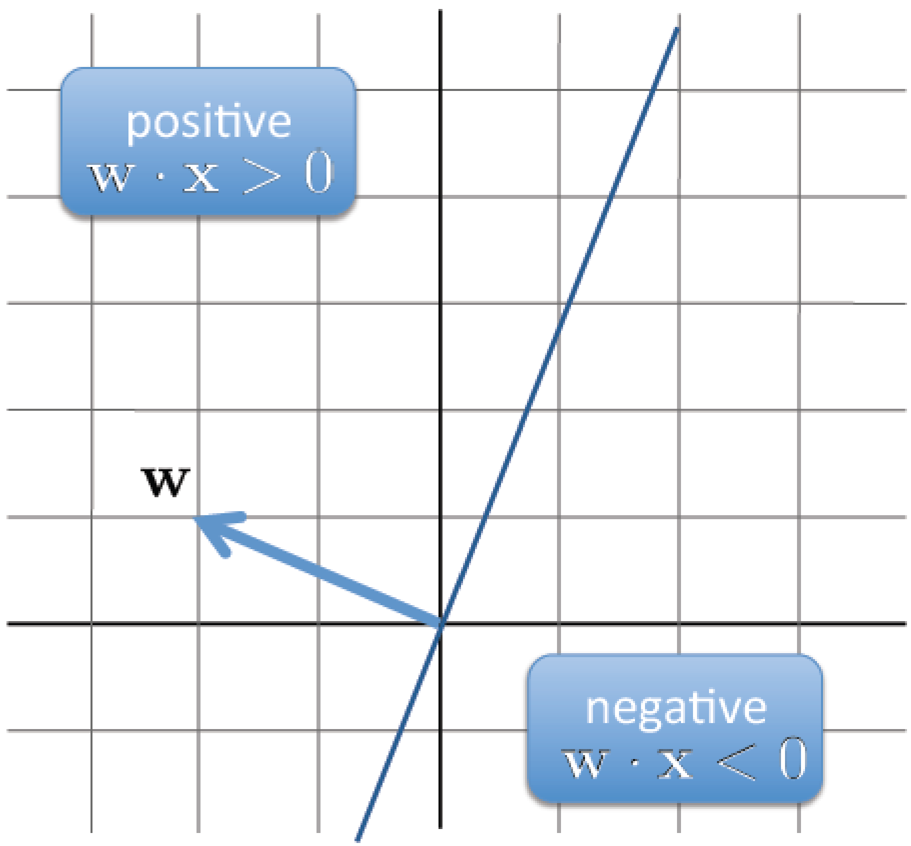
$$= y_i(\boldsymbol{w}^{(t)} \cdot \boldsymbol{x}_i) + \underline{\eta {y_i}^2 (\boldsymbol{x}_i \cdot \boldsymbol{x}_i)}$$

$$> y_i(\boldsymbol{w}^{(t)} \cdot \boldsymbol{x}_i)$$

$\mathbf{x}_i \neq \mathbf{0}$ならば
必ず0よりも大きい

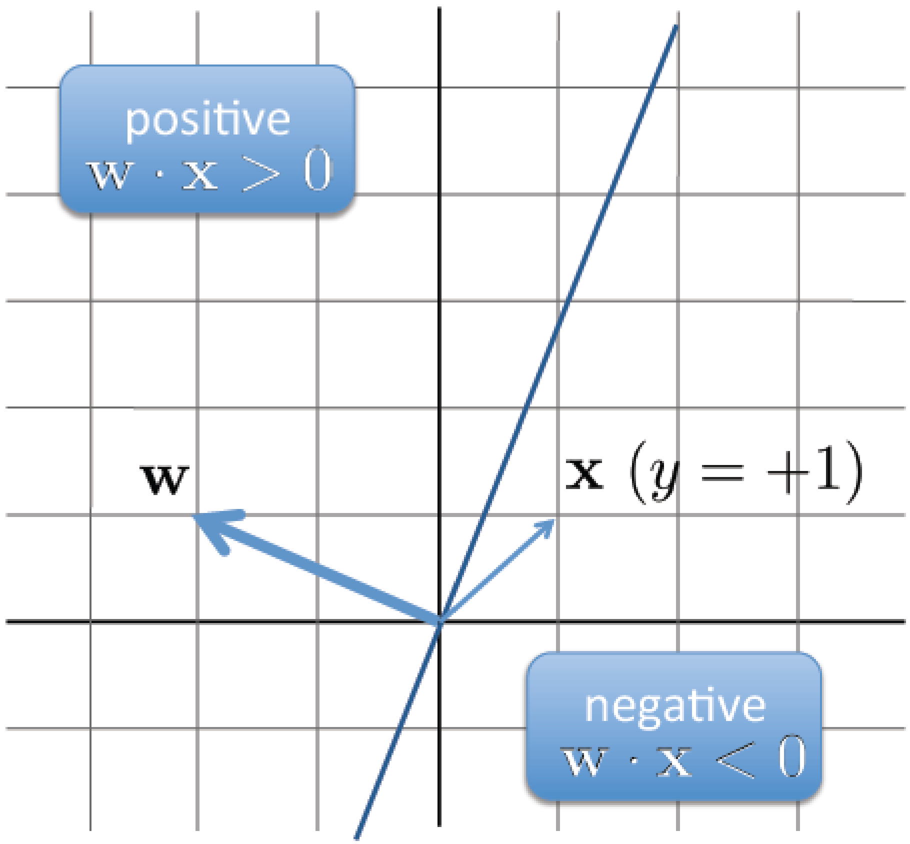
if文に引っかからないよう，条件式の中身が正の値に向かって更新されている

図で書くと(η=1のとき)



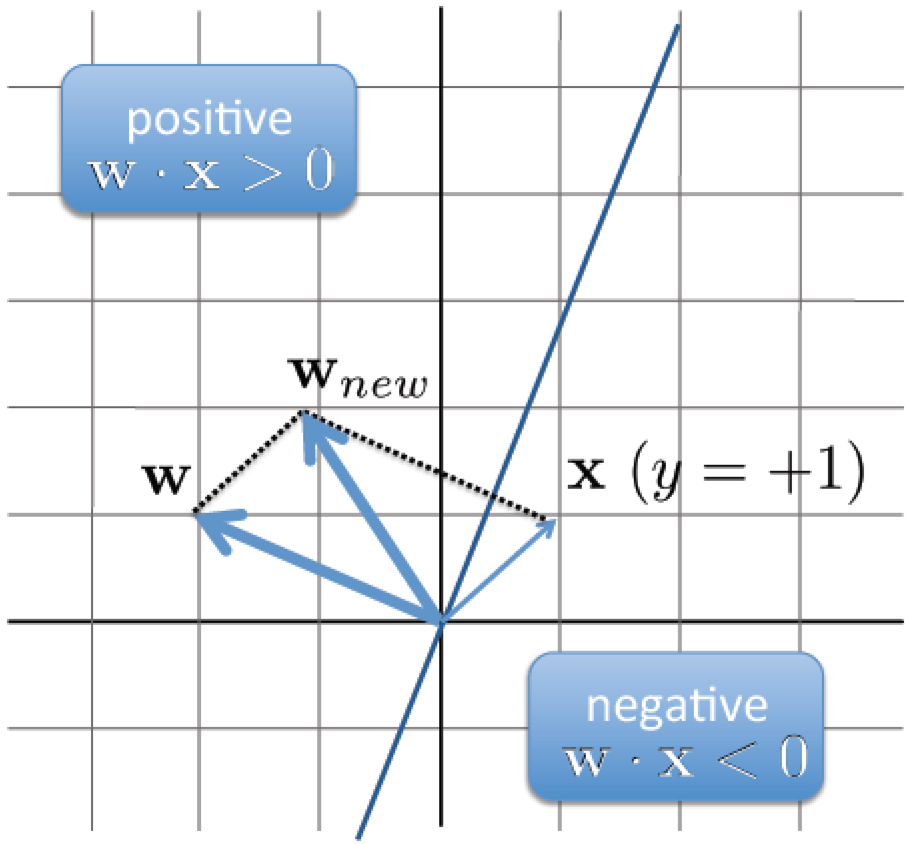
2008年度D-Lec 「DMLA Tutorial オンライン学習アルゴリズム（渡邉陽太郎）」より抜粋

図で書くと(η=1のとき)



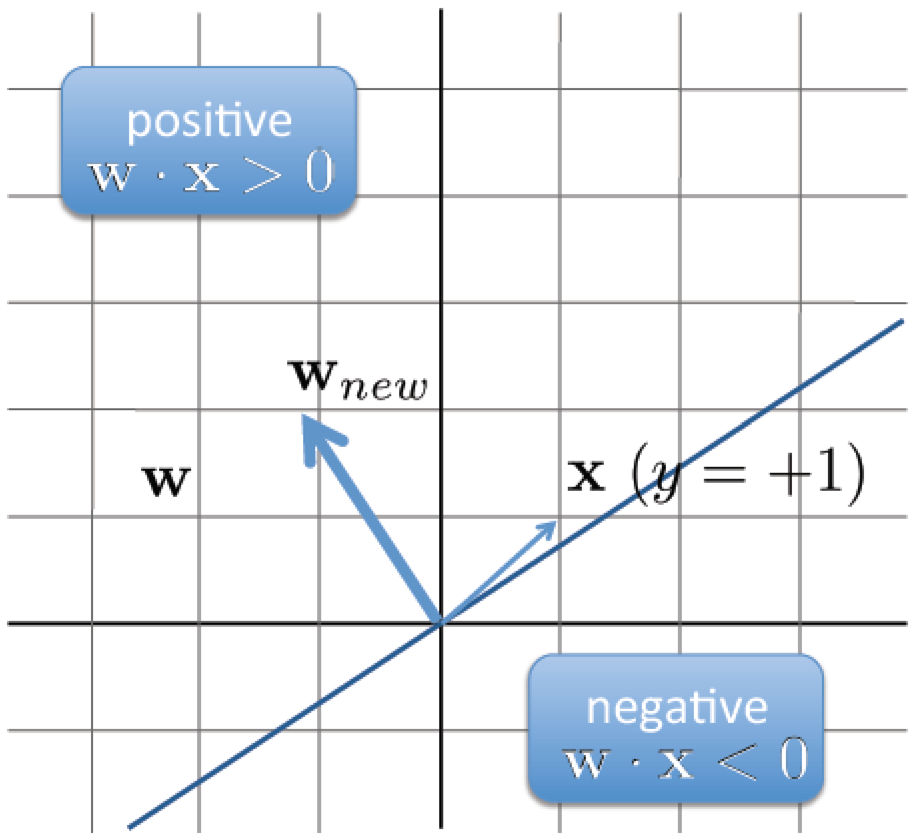
2008年度D-Lec 「DMLA Tutorial オンライン学習アルゴリズム(渡邉陽太郎)」より抜粋

図で書くと(η=1のとき)



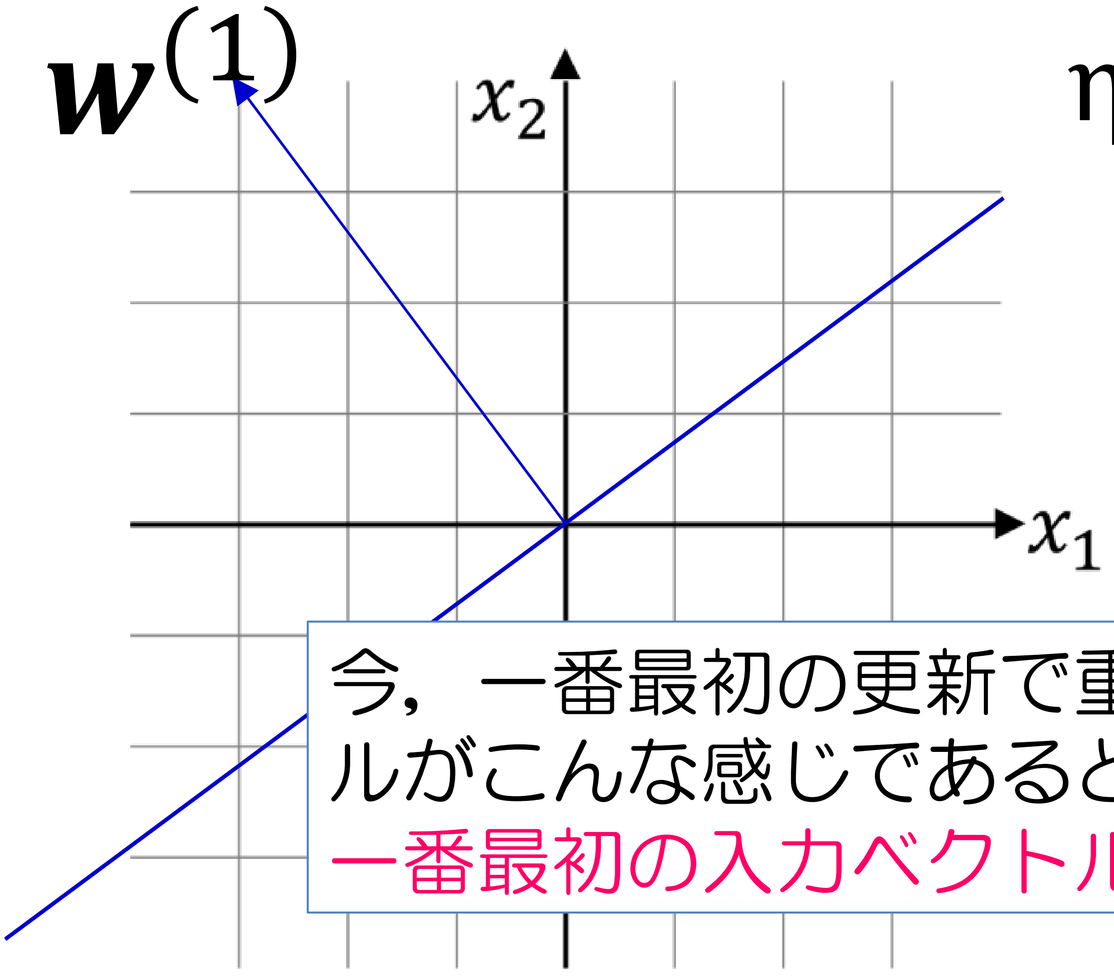
2008年度D-Lec 「DMLA Tutorial オンライン学習アルゴリズム(渡邉陽太郎)」より抜粋

図で書くと(η=1のとき)



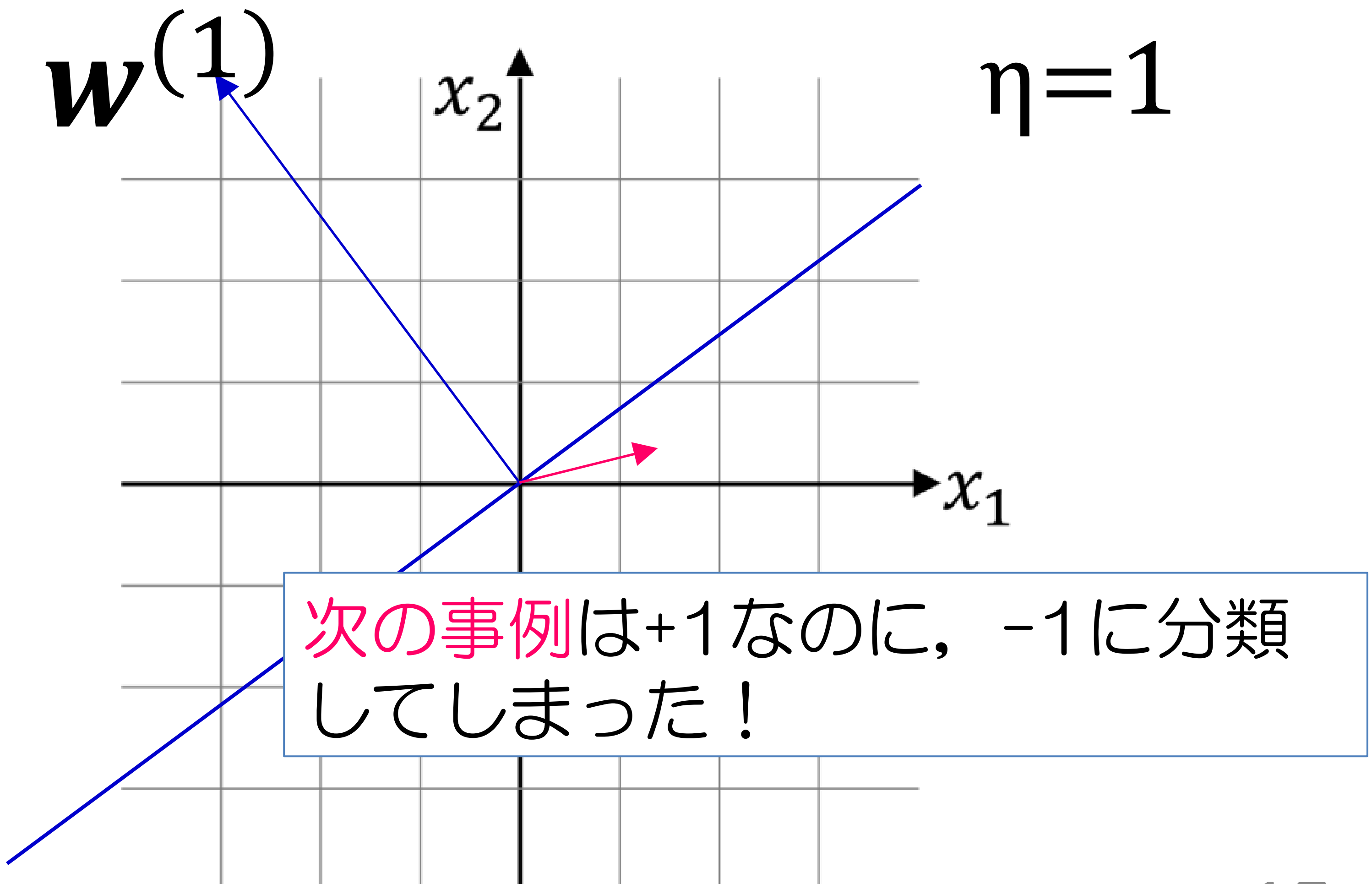
2008年度D-Lec 「DMLA Tutorial オンライン学習アルゴリズム(渡邉陽太郎)」より抜粋

ただし，１回の更新で必ず誤分類が直せるわけではない！

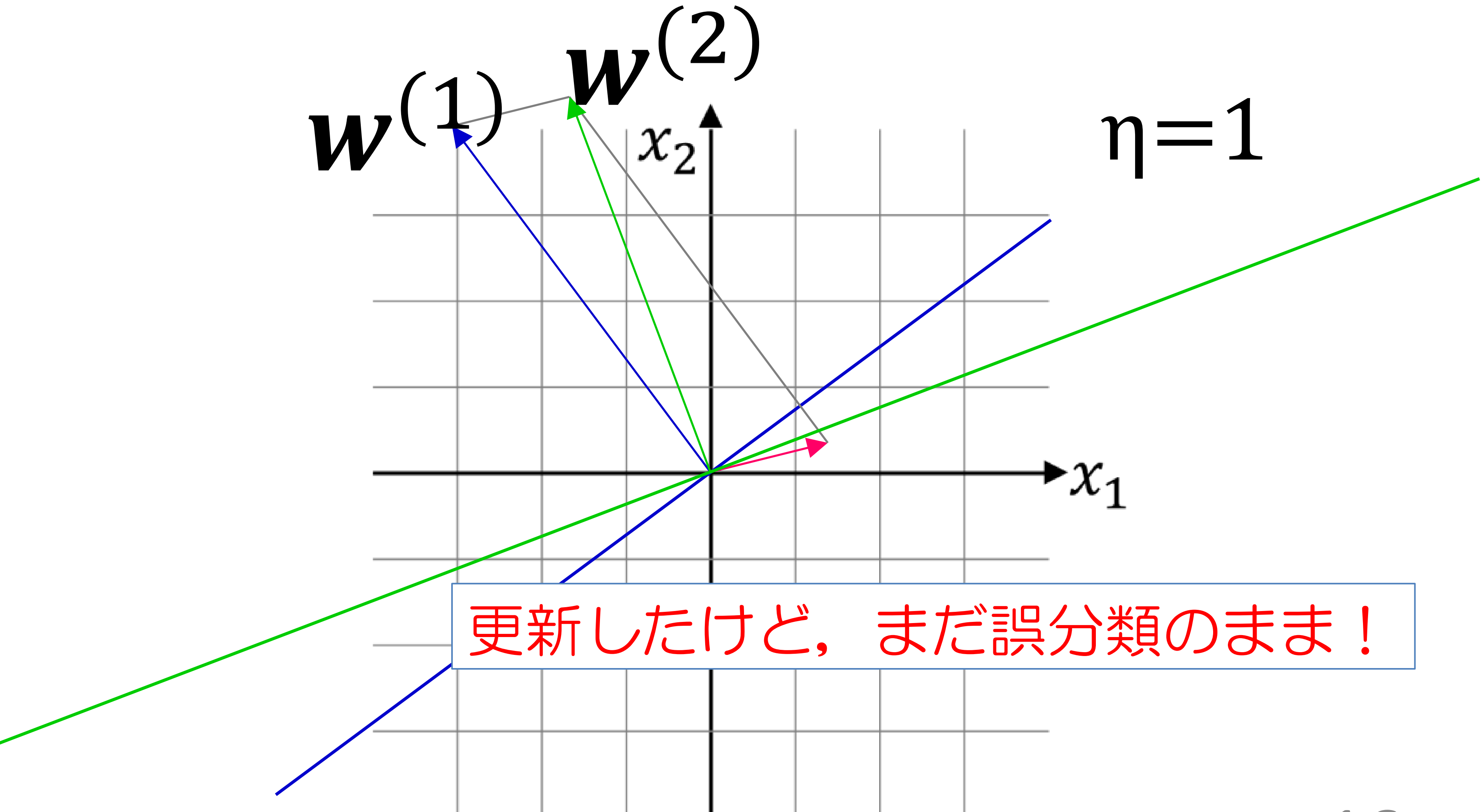


今，一番最初の更新で重みベクトルがこんな感じであるとする
一番最初の入力ベクトルと一致

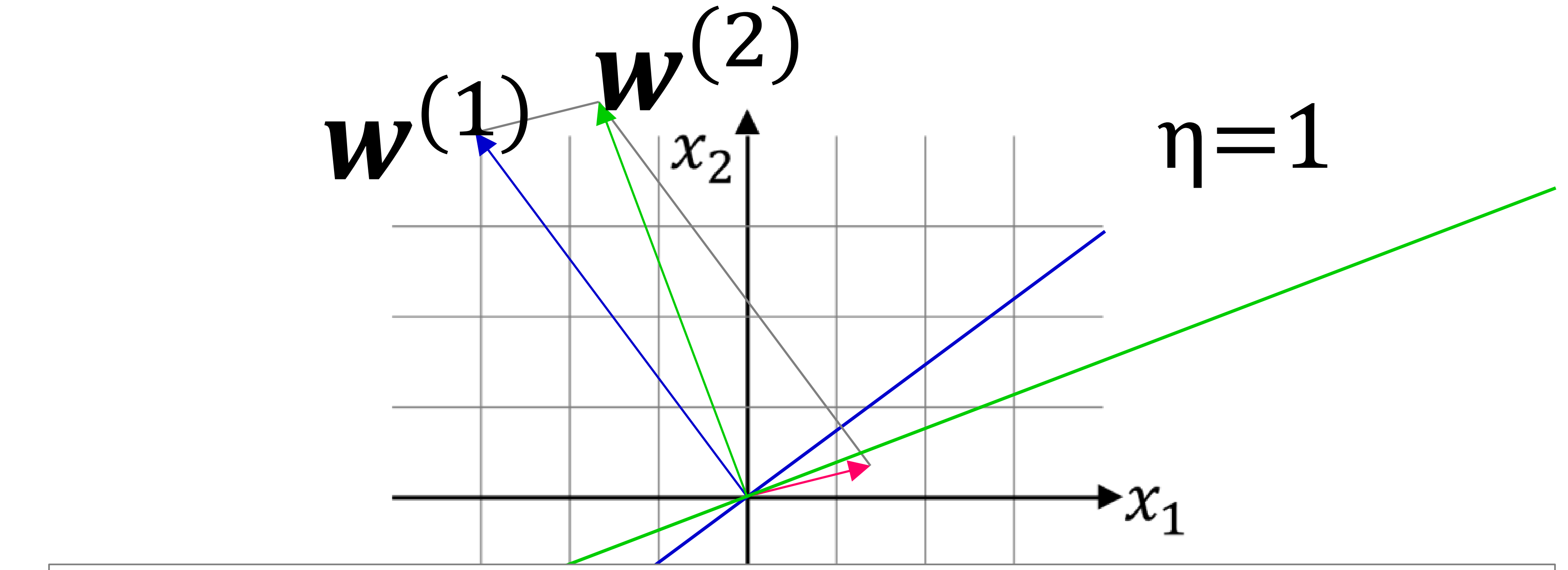
ただし，１回の更新で必ず誤分類が直せるわけではない！



ただし，１回の更新で必ず誤分類が直せるわけではない！

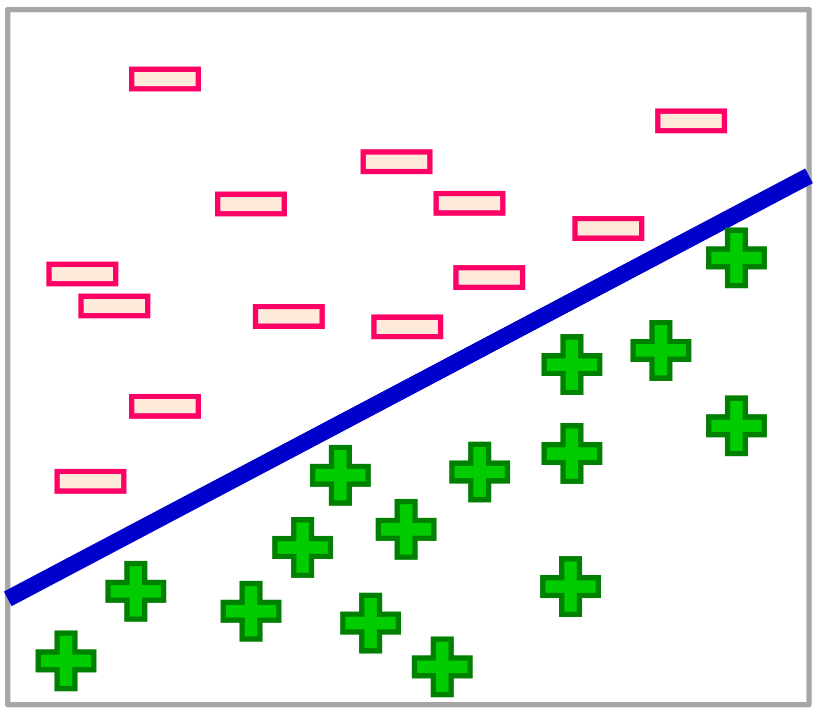


ただし，１回の更新で必ず誤分類が直せるわけではない！

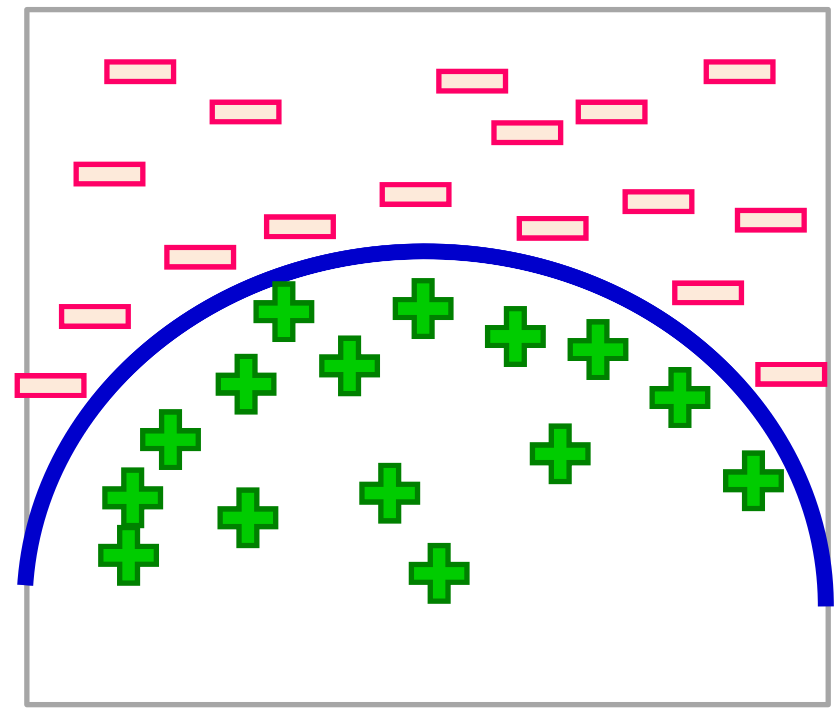


ポイント：
ベクトルの足し算の図形的意味から，
赤のベクトルの長さが，限りなく０に近づくと，緑のベクトルはどんどん青のベクトルに近づいていく．
＝　分離平面の回転角が小さくなる

コラム：　線形分離可能（1/2）



線形分離可能



線形分離不可能

コラム：　線形分離可能（2/2）

- 式で書くと…

$$|\boldsymbol{w}^*| = 1$$

$$\forall i \quad y_i(\boldsymbol{w}^* \cdot \boldsymbol{x}_i) > 0$$

- つまり，以下のような重みベクトル（分離平面）が存在することが線形分離可能の条件
  - 訓練事例を全て正しく分類
  - 分離平面上に訓練事例は１つも乗っていない
  - 長さ＝１
    - （重みベクトルの長さは分離平面の傾きとは無関係）

コラム：ノルムの２乗は自身の内積

$$|\boldsymbol{x}| = \sqrt{{x_1}^2 + {x_2}^2 + \cdots + {x_d}^2}$$

$$\boldsymbol{x} \in \mathbb{R}^d$$

より,

$$|\boldsymbol{x}|^2 = {x_1}^2 + {x_2}^2 + \cdots + {x_d}^2$$

$$= \boldsymbol{x} \cdot \boldsymbol{x}$$

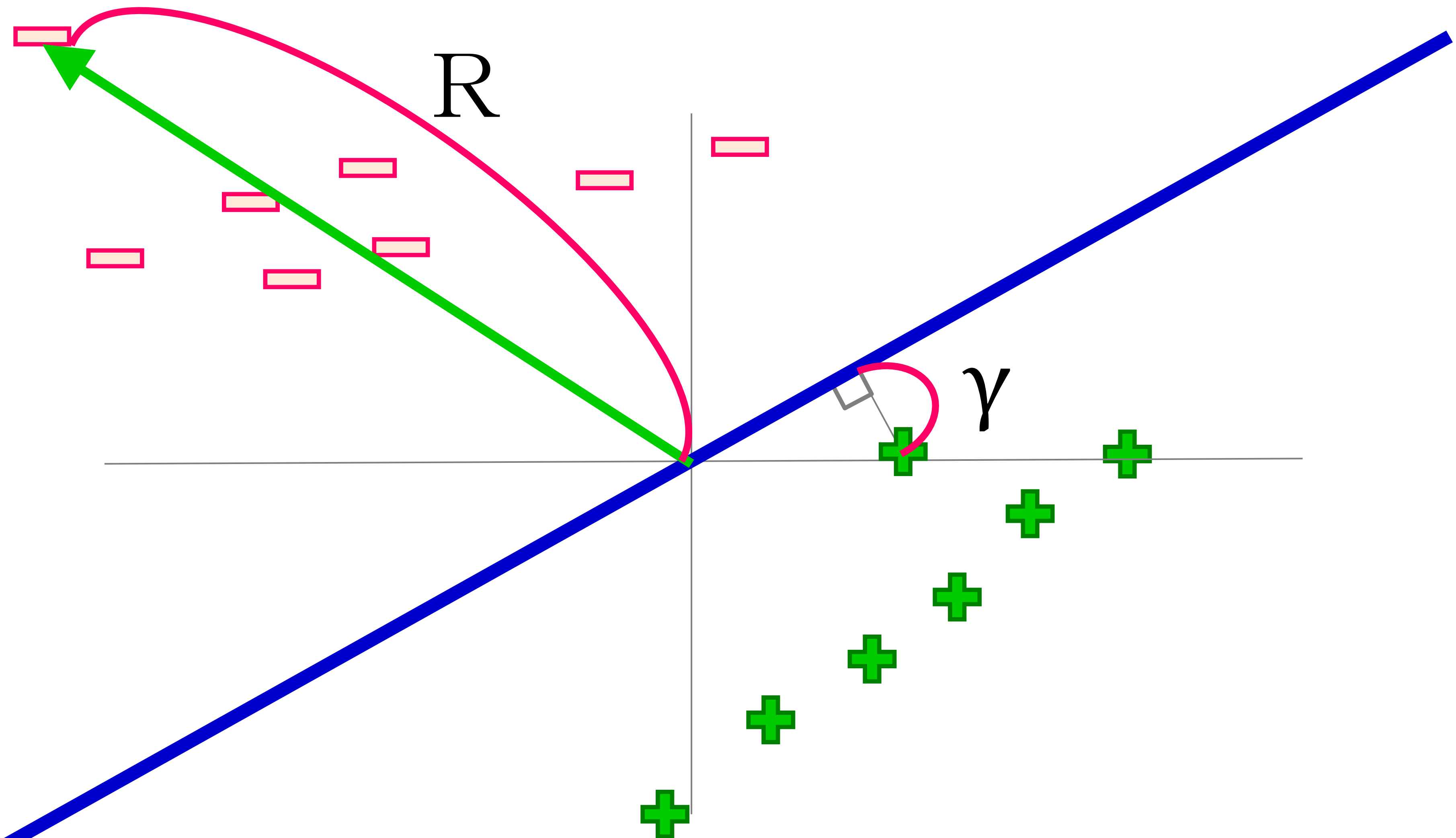
パーセプトロンの収束定理

- **訓練事例集合が線形分離可能ならば，以下の定理が成立する**
  - パーセプトロンは有限時間で学習を完了する

$$\text{重みベクトルの更新回数 } t \leq \frac{R^2}{\gamma^2}$$

R: 訓練事例中で最も長いベクトルの長さ
γ: 学習後に得られる分離平面から最近傍の訓練事例までの距離（マージン）

Rとγを図で書くと



証明（1/7）

- 学習の途中，i番目の事例(x_i, y_i)の分類に失敗し，重みベクトルのt回目の更新が行われたとする

$$\left|\boldsymbol{w}^{(t)}\right|^2 = \boldsymbol{w}^{(t)} \cdot \boldsymbol{w}^{(t)}$$

η>0

$$= (\boldsymbol{w}^{(t-1)} + \eta y_i \boldsymbol{x}_i) \cdot (\boldsymbol{w}^{(t-1)} + \eta y_i \boldsymbol{x}_i)$$

$$= (\boldsymbol{w}^{(t-1)} \cdot \boldsymbol{w}^{(t-1)}) + 2\eta y_i (\boldsymbol{w}^{(t-1)} \cdot \boldsymbol{x}_i) + \eta^2 {y_i}^2 (\boldsymbol{x}_i \cdot \boldsymbol{x}_i)$$

$$= \left|\boldsymbol{w}^{(t-1)}\right|^2 + \underline{2\eta y_i (\boldsymbol{w}^{(t-1)} \cdot \boldsymbol{x}_i)} + \eta^2 |\boldsymbol{x}_i|^2$$

≦0 (if文に引っかかっているから)

$$\leq \left|\boldsymbol{w}^{(t-1)}\right|^2 + \eta^2 |\boldsymbol{x}_i|^2$$

証明（2 /7 ）

$$\left|\boldsymbol{w}^{(t)}\right|^2 \leq \left|\boldsymbol{w}^{(t-1)}\right|^2 + \eta^2 |\boldsymbol{x}_i|^2$$

漸化式

$\boldsymbol{w}^{(0)} = \boldsymbol{0}$より

$$\left|\boldsymbol{w}^{(t)}\right|^2 \leq t\eta^2 R^2 \quad \cdots ①$$

R: 訓練事例中で最も長いベクトルの長さ

証明（3 /7）

- 学習の結果，すべての訓練事例を正しく分類し，かつ分離平面上に訓練事例の乗っていない重みベクトル　$\mathbf{w}^*$　が得られたとする
  - 線形分離可能なら，そういった$\mathbf{w}^*$が存在する
  - $\mathbf{w}^*$ は $|\mathbf{w}^*|=1$ になるようスケーリングされているとする
    - (分離平面の傾きはノルムの長さには依存しないので)

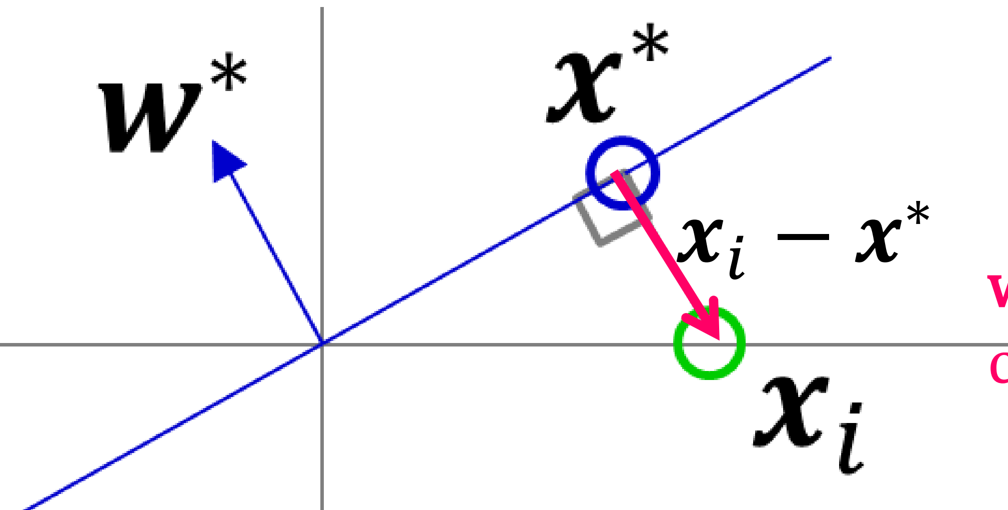
証明（4 / 7）

$$\boldsymbol{w}^* \cdot \boldsymbol{w}^{(t)} = \boldsymbol{w}^* \cdot (\boldsymbol{w}^{(t-1)} + \eta y_i \boldsymbol{x}_i)$$

$$= (\boldsymbol{w}^* \cdot \boldsymbol{w}^{(t-1)}) + \eta y_i (\boldsymbol{w}^* \cdot \boldsymbol{x}_i)$$

$$= (\boldsymbol{w}^* \cdot \boldsymbol{w}^{(t-1)}) + \eta y_i ((\boldsymbol{w}^* \cdot \boldsymbol{x}_i) - \overbrace{(\boldsymbol{w}^* \cdot \boldsymbol{x}^*)}^{=0})$$

$$= (\boldsymbol{w}^* \cdot \boldsymbol{w}^{(t-1)}) + \eta y_i (\boldsymbol{w}^* \cdot (\boldsymbol{x}_i - \boldsymbol{x}^*))$$



$\mathbf{w}^*$と$(\mathbf{x}_i$-$\mathbf{x}^*)$のなす角をθとすると
cosθと$\mathbf{y}_i$は符号が反対になる

$$= (\boldsymbol{w}^* \cdot \boldsymbol{w}^{(t-1)}) + \eta \, |\boldsymbol{w}^*||(\boldsymbol{x}_i - \boldsymbol{x}^*)|$$

証明（5 /7）

$$\boldsymbol{w}^* \cdot \boldsymbol{w}^{(t)} = (\boldsymbol{w}^* \cdot \boldsymbol{w}^{(t-1)}) + \eta \,\underbrace{|\boldsymbol{w}^*|}_{=1}|(\boldsymbol{x}_i - \boldsymbol{x}^*)|$$

$$= (\boldsymbol{w}^* \cdot \boldsymbol{w}^{(t-1)}) + \eta\,|(\boldsymbol{x}_i - \boldsymbol{x}^*)|$$

$$\geq (\boldsymbol{w}^* \cdot \boldsymbol{w}^{(t-1)}) + \eta\gamma$$

漸化式

γ：$\mathbf{w}^*$がなす分離平面から最近傍の訓練事例までの距離（マージン）

$\boldsymbol{w}^{(0)} = \mathbf{0}$より，$\boldsymbol{w}^* \cdot \boldsymbol{w}^{(0)} = \mathbf{0}$なので，

$$\boldsymbol{w}^* \cdot \boldsymbol{w}^{(t)} \geq t\eta\gamma \quad \cdots ②$$

証明（6 /7）

シュワルツの不等式：

$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ のとき，

$$(\boldsymbol{x} \cdot \boldsymbol{y})^2 \leq |\boldsymbol{x}|^2 |\boldsymbol{y}|^2$$

より，式①と式②を組み合わせて，

$$\left|\boldsymbol{w}^{(t)}\right|^2 \leq t\eta^2 R^2 \quad \cdots ①$$

$$\boldsymbol{w}^* \cdot \boldsymbol{w}^{(t)} \geq t\eta\gamma \quad \cdots ②$$

証明（7 /7）

$$t^2\eta^2\gamma^2 \le \left(\boldsymbol{w}^* \cdot \boldsymbol{w}^{(t)}\right)^2 \quad \cdots ②'$$

$$\le |\boldsymbol{w}^*|^2 \left|\boldsymbol{w}^{(t)}\right|^2$$

$$= \left|\boldsymbol{w}^{(t)}\right|^2 \le t\eta^2 R^2 \quad \cdots ①$$

$$t^2\eta^2\gamma^2 \le t\eta^2 R^2$$

$$t \le \frac{R^2}{\gamma^2}$$

ポイント：
学習率ηは収束とは無関係
つまり，０よりも大きな値ならばなんでもいい

収束まとめ

シュワルツの不等式により

いつかは止まる