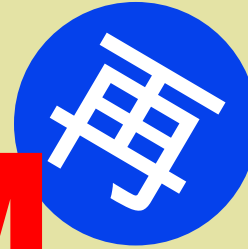


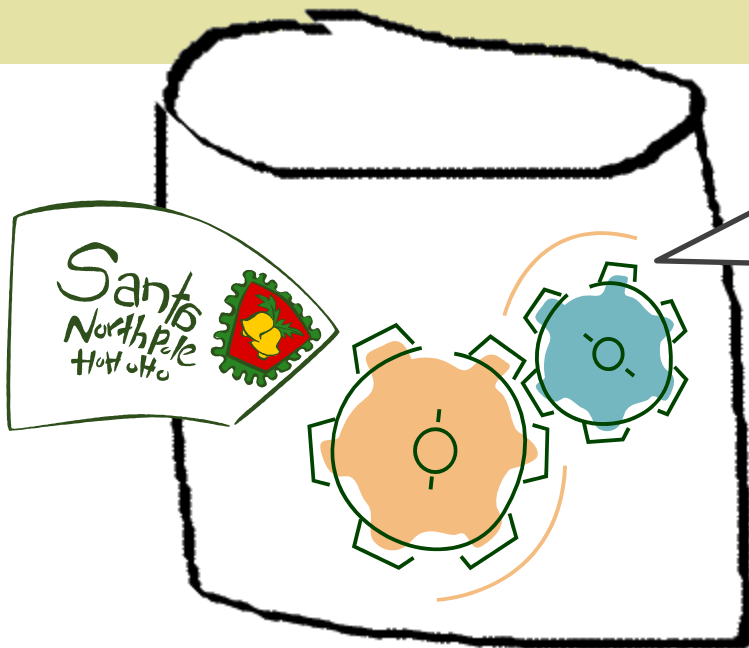
これなら

わかるSVM



挫折しない
分類器入門

@D-Lec



スパムだ！

スパムだ！

ここにはある学生の顔写真が入っていましたが、肖像権の関係で削除しました。

本日の見どころ

- タスクを 2 値分類問題として定式化する

- 分離平面

- ヒンジロス

- マージン最大化

✗ 学習（最適化）  来週, TBKくんがやってくれるはず

SVMとは…

- 「予め設定した次元数(**d**)」の実数値ベクトルを入力すると「**正**」か「**負**」を返す関数

- 例：

$$x_1, x_2 \in \mathbb{R}^d$$

$$svm(x_1) = \text{正}^{(\text{正例})}$$

$$svm(x_2) = \text{負}^{(\text{負例})}$$

2値分類

もっと詳しくいうと…

- 実際は，入力されたベクトルに対して， $-\infty \sim +\infty$ の実数値をスコアとして返す関数があって，SVMはそのスコアの符号を返す

$$svm(x) = \begin{cases} \text{正} & score(x) \geq 0 \\ \text{負} & score(x) < 0 \end{cases}$$

例：
 $score(x_1) = 13.21333 \Rightarrow svm(x_1) = \text{正}$
 $score(x_2) = -0.0026 \Rightarrow svm(x_2) = \text{負}$

じゃあ、スコアって？

- スコア関数は以下の式で表される

$$\text{score}(x) = \mathbf{w} \cdot \mathbf{x}$$

重みベクトル
↓
ベクトルの内積

$$\mathbf{w}, \mathbf{x} \in \mathbb{R}^d$$

ここに $-b$ （バイアス項）とか付くこともあるけど簡単のため省略.

まとめると

$$svm(\boldsymbol{x}) = \begin{cases} \text{正} & score(\boldsymbol{x}) \geq 0 \\ \text{負} & score(\boldsymbol{x}) < 0 \end{cases}$$

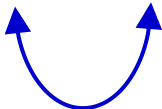
$$score(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x}$$

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

コラム1 ベクトルの内積(1/2)

- 2つの次元数が同じベクトル x_1, x_2 が与えられたとき, x_1 と x_2 の内積は以下のように計算できる

$$x_1 = \begin{bmatrix} x_{11} \\ \vdots \\ x_{1d} \end{bmatrix} \quad x_2 = \begin{bmatrix} x_{21} \\ \vdots \\ x_{2d} \end{bmatrix} \quad \text{次元数: } d$$

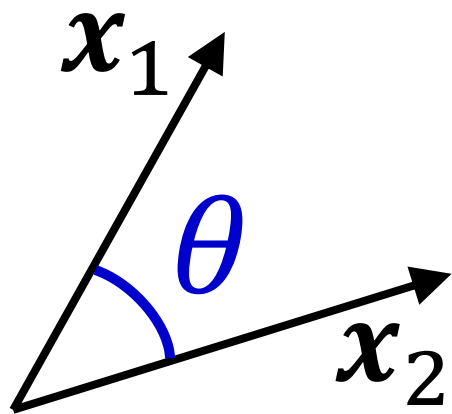
$$x_1 \cdot x_2 = (x_{11} \times x_{21}) + (x_{12} \times x_{22}) + \cdots + (x_{1d} \times x_{2d})$$


左右入れ替えても結果は同じ値

コラム1 ベクトルの内積(2/2)

□ 内積は以下のようにも計算可能

□ 後で出てきます！



θ : 2つのベクトルがなす角度

L2ノルム (ベクトルの長さ)

$$|x| = \sqrt{x_1^2 + x_2^2 + \dots + x_d^2} \\ \geq 0$$

$$x_1 \cdot x_2 = |x_1| |x_2| \cos \theta$$

左右入れ替えても結果は同じ値

じゃあ、NLPでどう使うの？

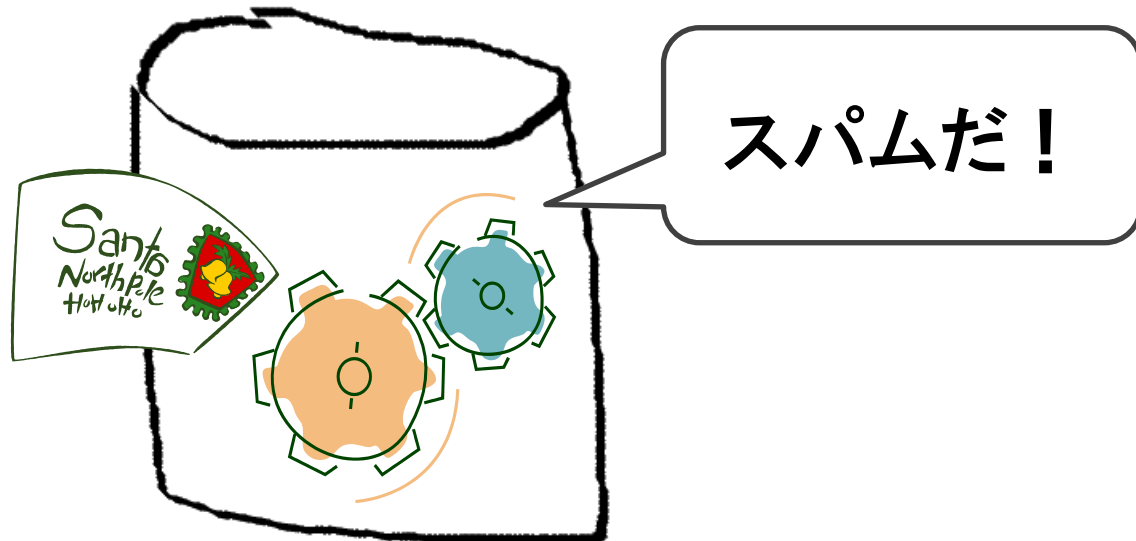
事例

- ① 分類したいもの(e.g., 文書, 文, 単語, 文字列 etc...)をベクトルで表現
- ② SVMで2値分類
- ③ 2値分類した結果(正or負)を分類先のラベルに書き換え

↑
クラスラベル

例：スパム判定

- 事例： メール（1通）
- クラストラベル：
 - スпамだ！
 - スпамじゃない！



素性ベクトルと素性関数

重要

□ 事例をベクトルで表現

- 方法：メールをベクトルに変換する関数を作成

$$\phi(\text{メール}) = \begin{bmatrix} \phi_1(\text{メール}) \\ \phi_2(\text{メール}) \end{bmatrix}$$

素性ベクトル

素性関数

$$\phi_1(\text{メール}) = \begin{cases} 1 & \text{メール文内に単語「儲かる」が入っている} \\ 0 & \text{入っていない} \end{cases}$$

$$\phi_2(\text{メール}) = \begin{cases} 1 & \text{メール文内に単語「NLP」が入っている} \\ 0 & \text{入っていない} \end{cases}$$

正負の値とクラスラベル

□ SVMの出力（正or負）をクラスラベルに対応付ける

- 正： スпамだ！
- 負： スпамじゃない！

ここまでの流れを指して、
「スパム判定を2値分類問題として定式化した」
という

実際に分類してみよう

メール1

楽しく儲かる
方法教えます！

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

スパム

メール2

NLP勉強会を
開催します。

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

スパム
じゃない

メール3

NLPで儲かる
20の秘訣！

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

スパム

重みベクトル w は適当に設定

$$w = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

まず、メール1

メール1

楽しく儲かる
方法教えます！

スパム

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

$$\begin{aligned} \text{score} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) &= \begin{bmatrix} 1 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= (1 \times 1) + (-2 \times 0) \\ &= 1 \geq 0 \end{aligned}$$

よって、SVMの出力は正
つまり、スパムがちゃんと
「スパムである」と判定された！

次に、メール2

メール2

NLP勉強会を
開催します。


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

$$\begin{aligned} \text{score} \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) &= \begin{bmatrix} 1 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= (1 \times 0) + (-2 \times 1) \\ &= -2 < 0 \end{aligned}$$

よって、SVMの出力は負
つまり、スパムでないメールがちゃんと
「スパムではない」と判定された！

最後に、メール3

メール3

スパム

NLPで儲かる
20の秘訣！

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

$$\begin{aligned} \text{score} \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) &= \begin{bmatrix} 1 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= (1 \times 1) + (-2 \times 1) \\ &= -1 < 0 \end{aligned}$$

判定失敗！

よって、SVMの出力は負
つまり、スパムが

「スパムではない」と判定されてしまった！16

重みを変えてみる

$$w = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad \rightarrow \quad w' = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

$$\begin{aligned} \text{score} \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) &= \begin{bmatrix} 3 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= (3 \times 1) + (-2 \times 1) \\ &= 1 \geq 0 \end{aligned}$$

判定成功！

よって、SVMの出力は正
つまり、スパムがちゃんと
「スパムである」と判定された！

これがつまり，学習

$$w = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \rightarrow w' = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

□ SVMの学習：

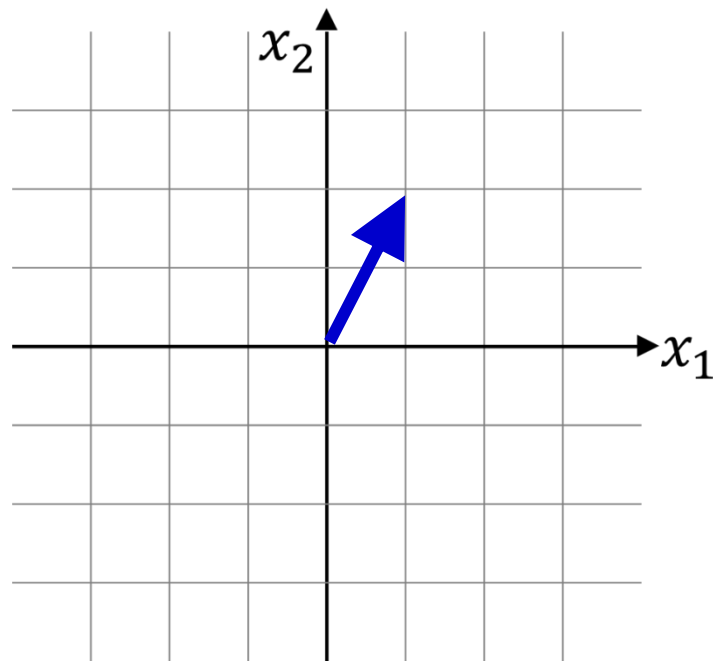
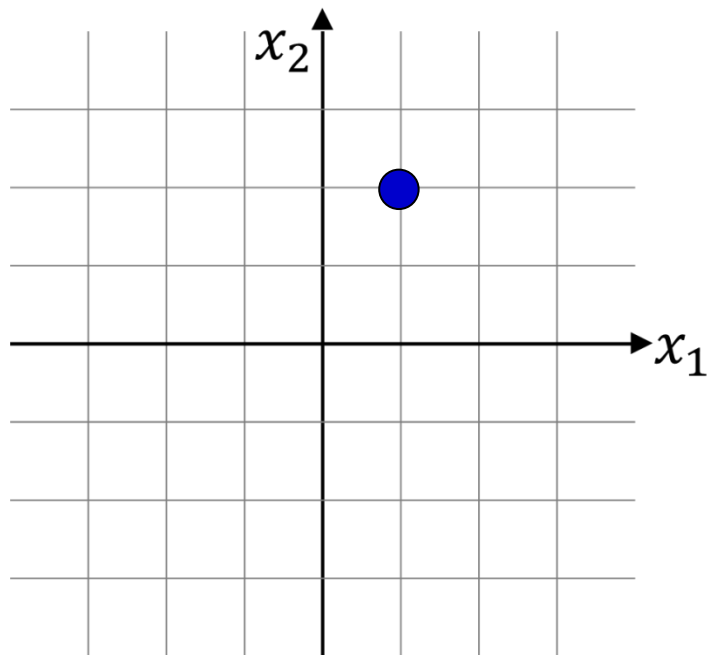
- 人手でラベル付された事例(訓練事例)が複数与えられたときに，それらを正しく分類できるようにパラメータ w を求める手続き

☞ 教師あり学習

- ちなみに，上の w' はメール3だけでなく，1と2も正しく分類可能

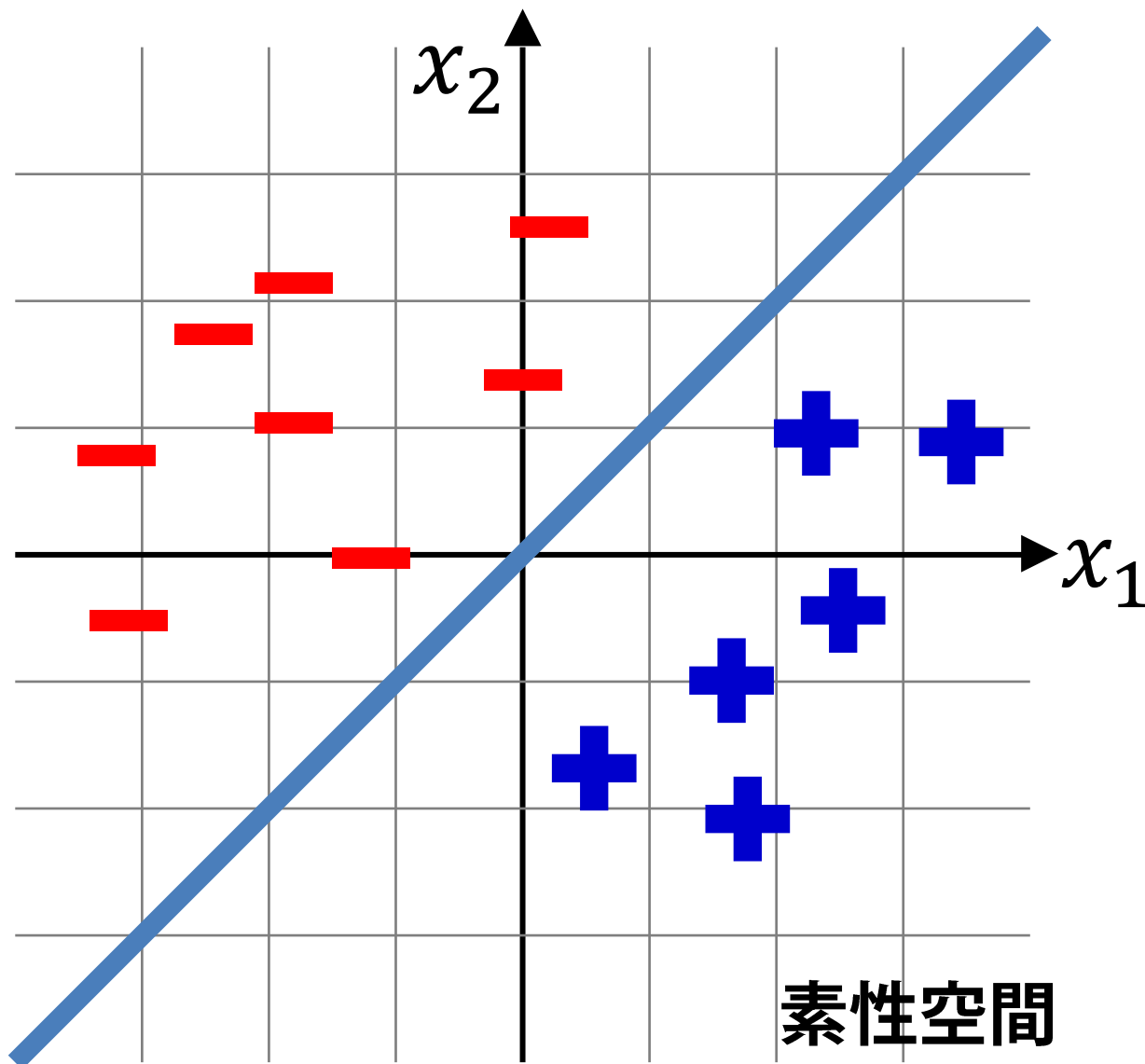
コラム2 ベクトルの表し方

例： $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$



分類器の話をするときは1つの図内で上記二つの表記が混在することもあるので注意

ところで、分離平面ってよく聞くよね？



+: 正例
-: 負例

分離平面の正体

□ 分離平面：

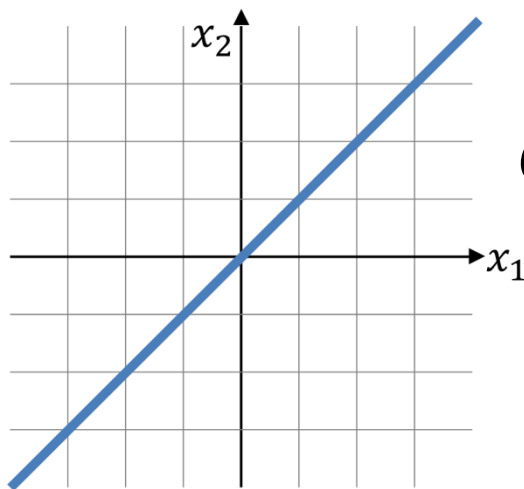
□ 重みベクトル \mathbf{w} がある値に定まったとき,

$$\text{score}(\mathbf{x}^*) = \mathbf{w} \cdot \mathbf{x}^* = 0$$

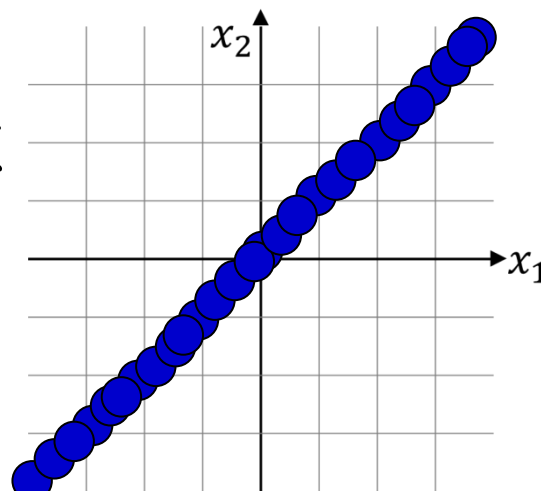
が成り立つようなベクトル \mathbf{x}^* の集合

$$\{\mathbf{x}^* \mid \mathbf{w} \cdot \mathbf{x}^* = 0, \mathbf{x}^*, \mathbf{w} \in \mathbb{R}^d\}$$

つまり,



の正体は



分離平面は重みベクトルと直交する

□ 証明

$w \neq 0$ とする

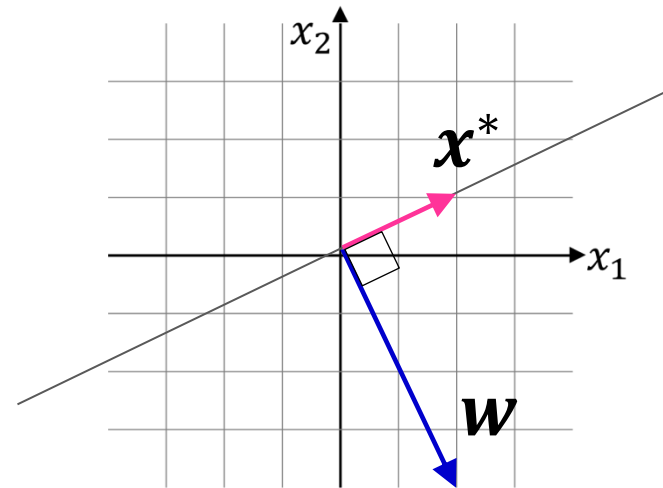
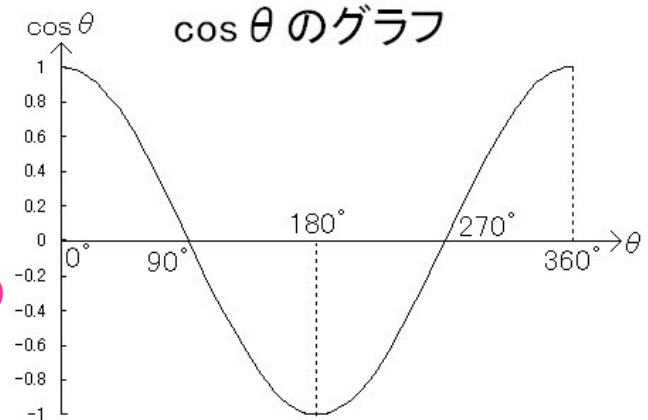
固定 $\rightarrow w \cdot x^* = 0$ ← 0以外の分離平面上の
任意のベクトル

$$\underset{0}{|w|} \underset{0}{|x^*|} \cos \theta = 0$$

$$\cos \theta = 0$$

$$\theta = 90^\circ$$

よって、 w は分離平面上のすべてのベクトルに対して垂直
なので、 w は分離平面と直交する



実は分離平面の定義から一発

$$\{x^* \mid \underbrace{w \cdot x^*}_{\text{2つのベクトルが直交する条件}} = 0, x, w \in \mathbb{R}^d\}$$

2つのベクトルが直交する条件

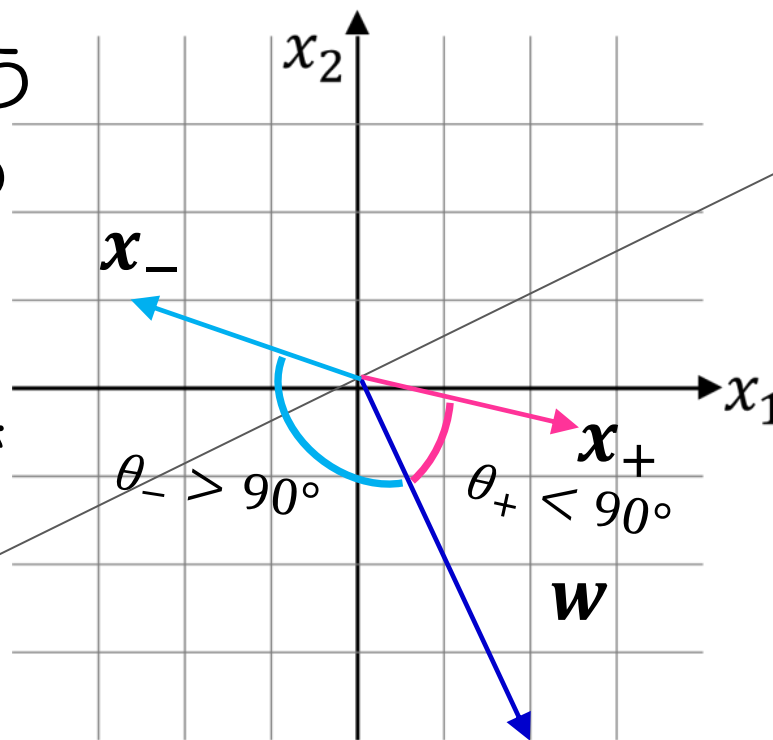
分離平面のw側はスコアが必ず正 反対は負

□ 証明：

- w と x^* がなす角が 90° ということは，分離平面の w 側にあるあらゆるベクトル x_+ は w となす角が 90° よりも小さい
- 反対に w と分離平面を挟んで反対にあるベクトル x_- が w となす角は 90° よりも大きい

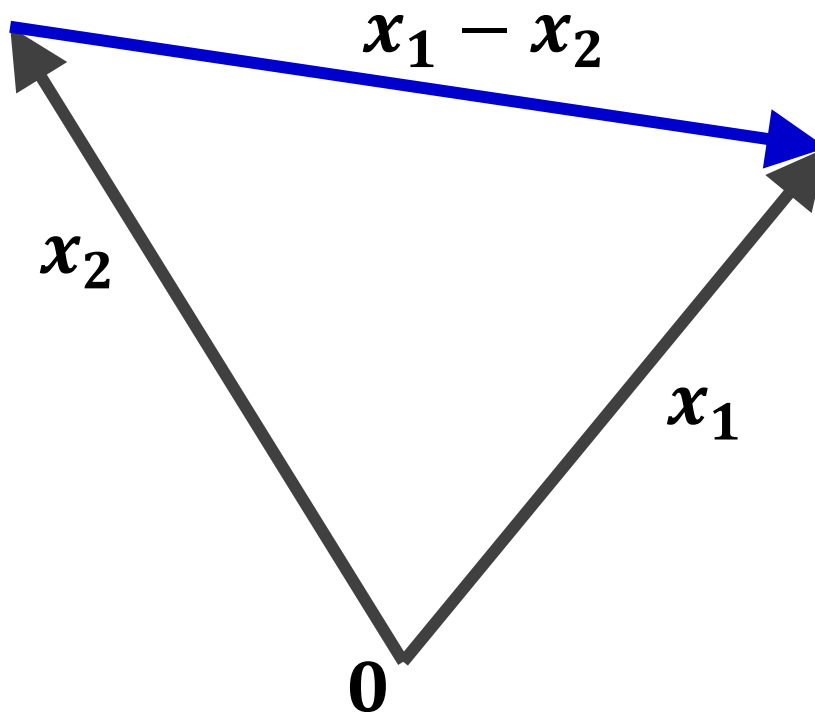
$$\underset{\substack{\downarrow \\ 0}}{\text{score}}(\underset{\substack{\downarrow \\ 0}}{x}) = \underset{\substack{\downarrow \\ 0}}{|w|} \underset{\substack{\downarrow \\ 0}}{|x|} \cos \theta$$

より，スコアの正負は θ の値によって決まるので，
表題が成り立つ



コラム3 ベクトルの差

- 2つのベクトルの差 $x_1 - x_2$ は以下のような図形的意味を持つ

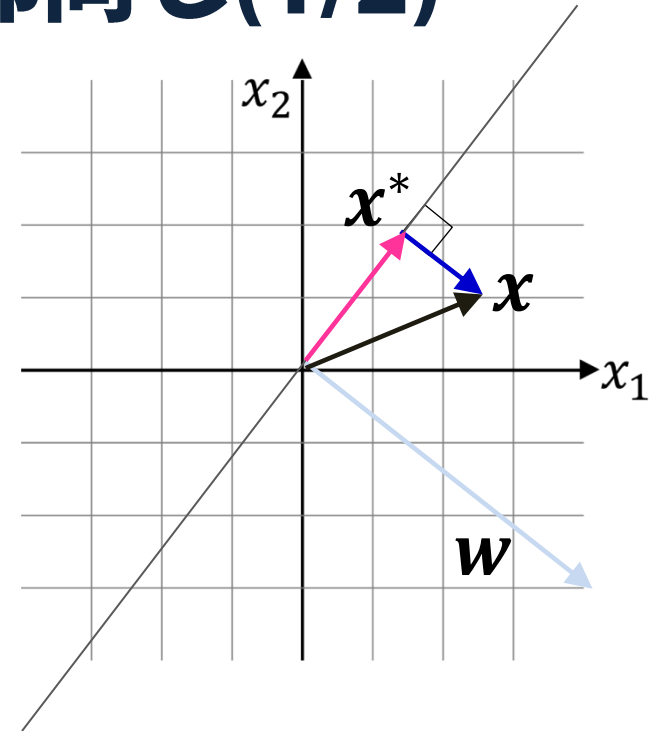


分離平面から等距離にあるベクトルのスコアの絶対値は全部同じ(1/2)

□ 証明

$$\begin{aligned} \text{score}(x) &= w \cdot x \\ &= w \cdot x - \overset{=0}{w \cdot x^*} \\ &= w \cdot (x - x^*) \\ &= |w| |x - x^*| \cos \theta \end{aligned}$$

ここで x^* を右のようなベクトルとする



すると、 w とベクトル $(x - x^*)$ がなす角は
 0° (正) or 180° (負)なので、 $\cos \theta = 1$ or -1

正負の区別を取るために絶対値を取ると


$$|\text{score}(x)| = |w| |x - x^*|$$

次ページへ→

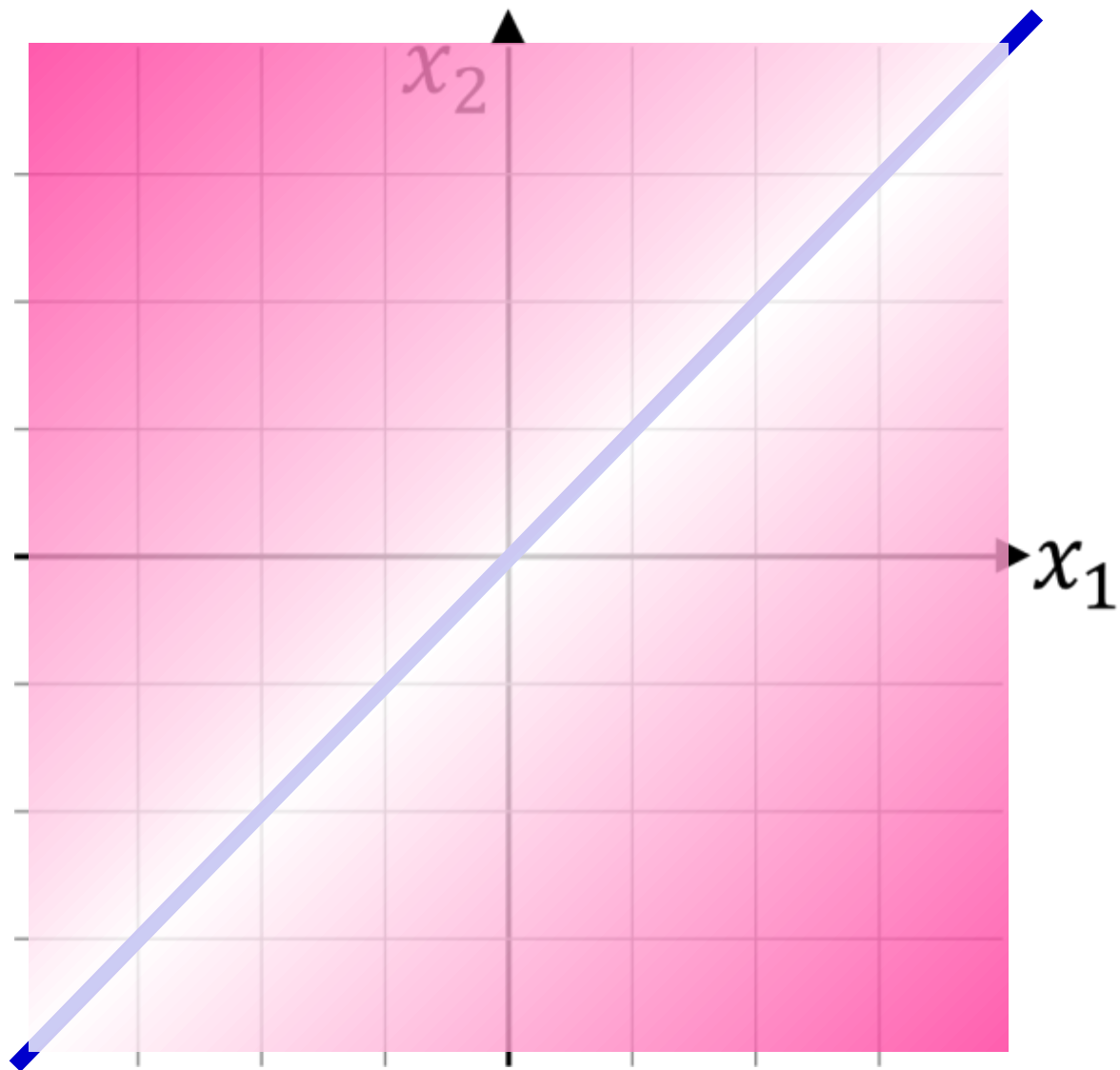
分離平面から等距離にあるベクトルの スコアの絶対値は全部同じ(2/2)

$$|score(x)| = |w||x - x^*|$$

ここで w を一定とすると, $score$ の大きさは $x - x^*$ の長さ(=分離平面からの直線距離)によって決まる

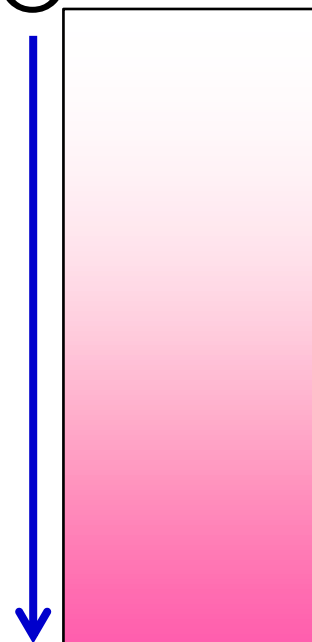
 分離平面から遠ざかるほどスコアの絶対値は高くなる

つまり, 分離平面からの直線距離が等しいベクトルのスコアの絶対値はすべて同じ



スコアの絶対値

0



高い

重みベクトルの大きさ $|w|$ は 分離平面の傾きとは無関係

□ 証明

$w_1 \cdot x^* = 0$ となるような x^* を考える

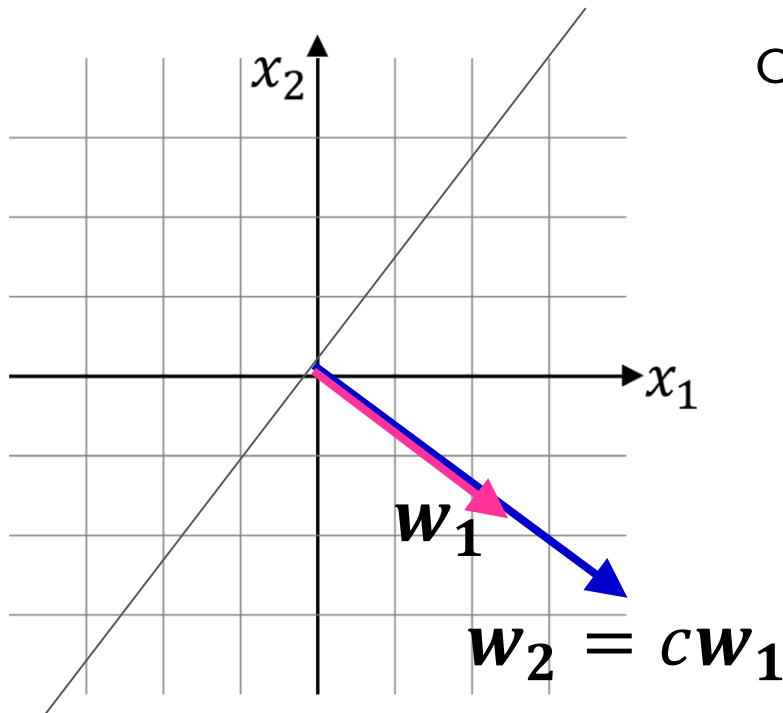
$w_2 = cw_1$ とおく (c は 0 よりも大きい定数)

$$w_2 \cdot x^* = cw_1 \cdot x^* = c(w_1 \cdot x^*) = 0$$

$c > 0$ より,

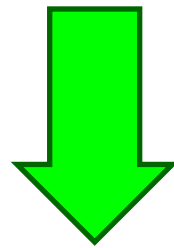
$$w_1 \cdot x^* = w_2 \cdot x^* = 0$$

つまり, $w_1 \cdot x^* = 0$ となる
に x^* においては $w_2 \cdot x^* = 0$ と
なり, また逆も成り立つた
ため, w_1 の分離平面と w_2 の
分離平面は同一



じつは、ここまでの話は、
パーセプトロンやPAでも同じ話

なら、SVMは何か違うのか？



最適な w を探す基準が違う！
(学習の基準)

SVMの損失関数

- SVMでは以下の値を最小にするような重みベクトル w を学習で求める

$$\max(0, \lambda - \underbrace{y\mathbf{w} \cdot \mathbf{x}}_{\text{スコア}})$$

y : 事例の正解ラベル

λ : 0以上の任意の定数
(1がよく使用される)

損失関数を読む

□ もし、事例 x の分類に失敗していたら…

□ y と $w \cdot x$ の符号は反対 ($-yw \cdot x \geq 0$)

➡ $\max(0, \lambda - yw \cdot x) = \lambda - yw \cdot x$

□ もし、事例 x の分類に成功していたら…

□ y と $w \cdot x$ の符号は一致 ($-yw \cdot x \leq 0$)

□ もし、スコアの絶対値が λ 未満なら…

□ $0 < \lambda - yw \cdot x \leq \lambda$

➡ $\max(0, \lambda - yw \cdot x) = \lambda - yw \cdot x$

□ もし、スコアの絶対値が λ 以上なら…

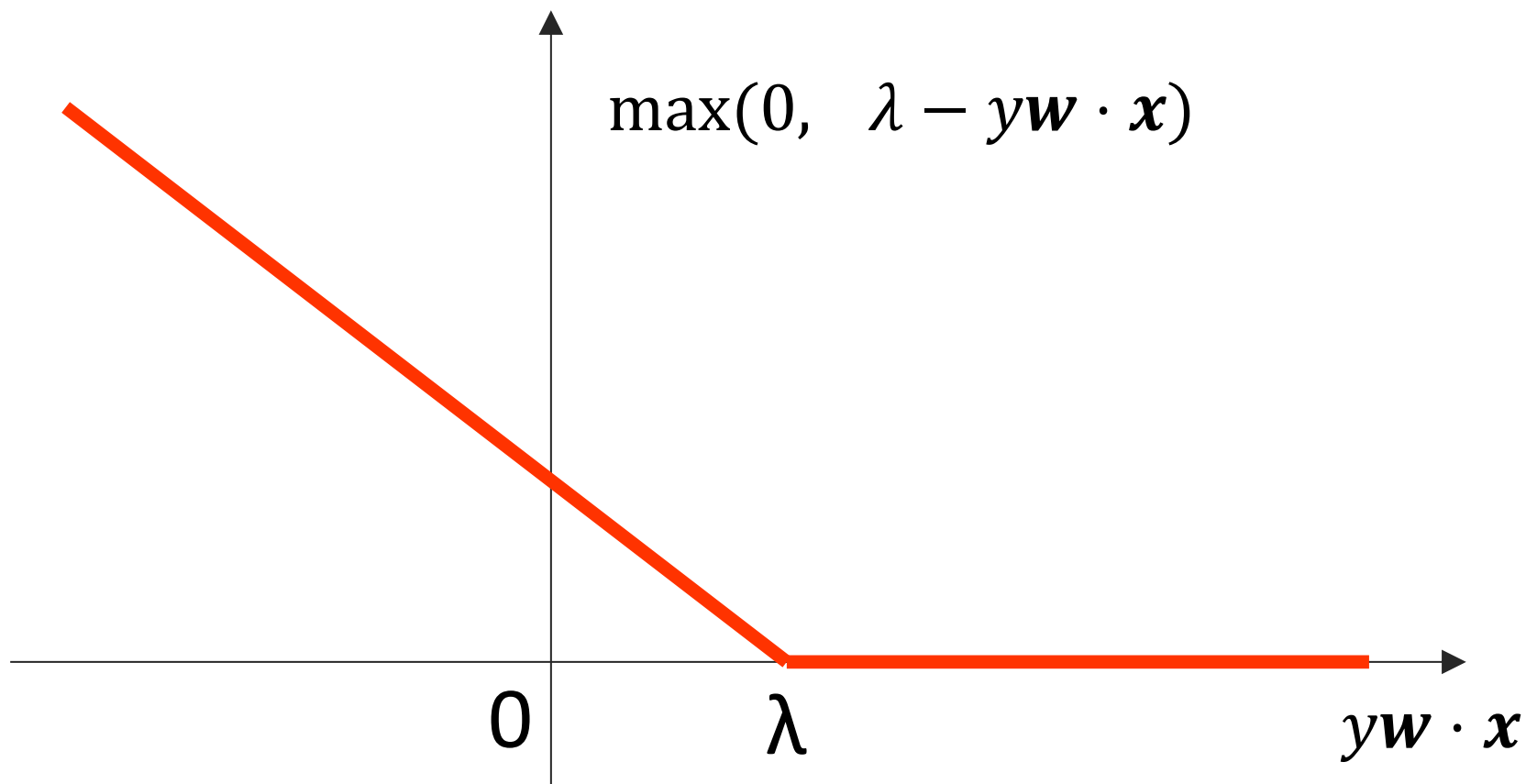
□ $\lambda - yw \cdot x \leq 0$

➡ $\max(0, \lambda - yw \cdot x) = 0$

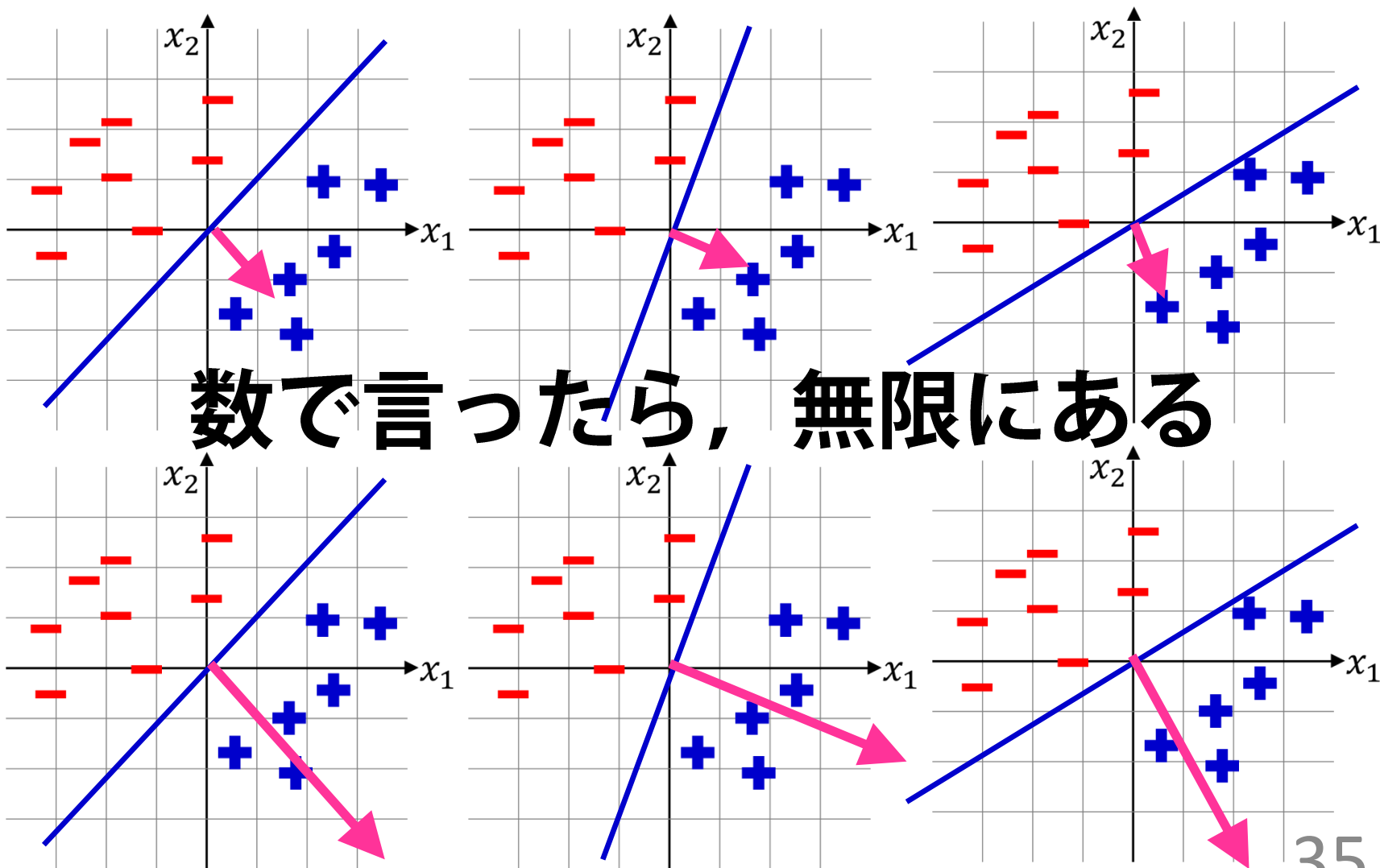
損失関数の意味

- スコアの絶対値が λ 以上で分類に成功していれば損失は 0 だが,
- 分類に失敗していたり,
- 分類に成功していてもスコアの絶対値が λ 未満ならば,
- ペナルティが付く.
- SVMはこのペナルティを最小にする重みベクトルを学習で探す.

ヒンジロス



損失を最小にするwはたくさん



そういった w を集合で表すと

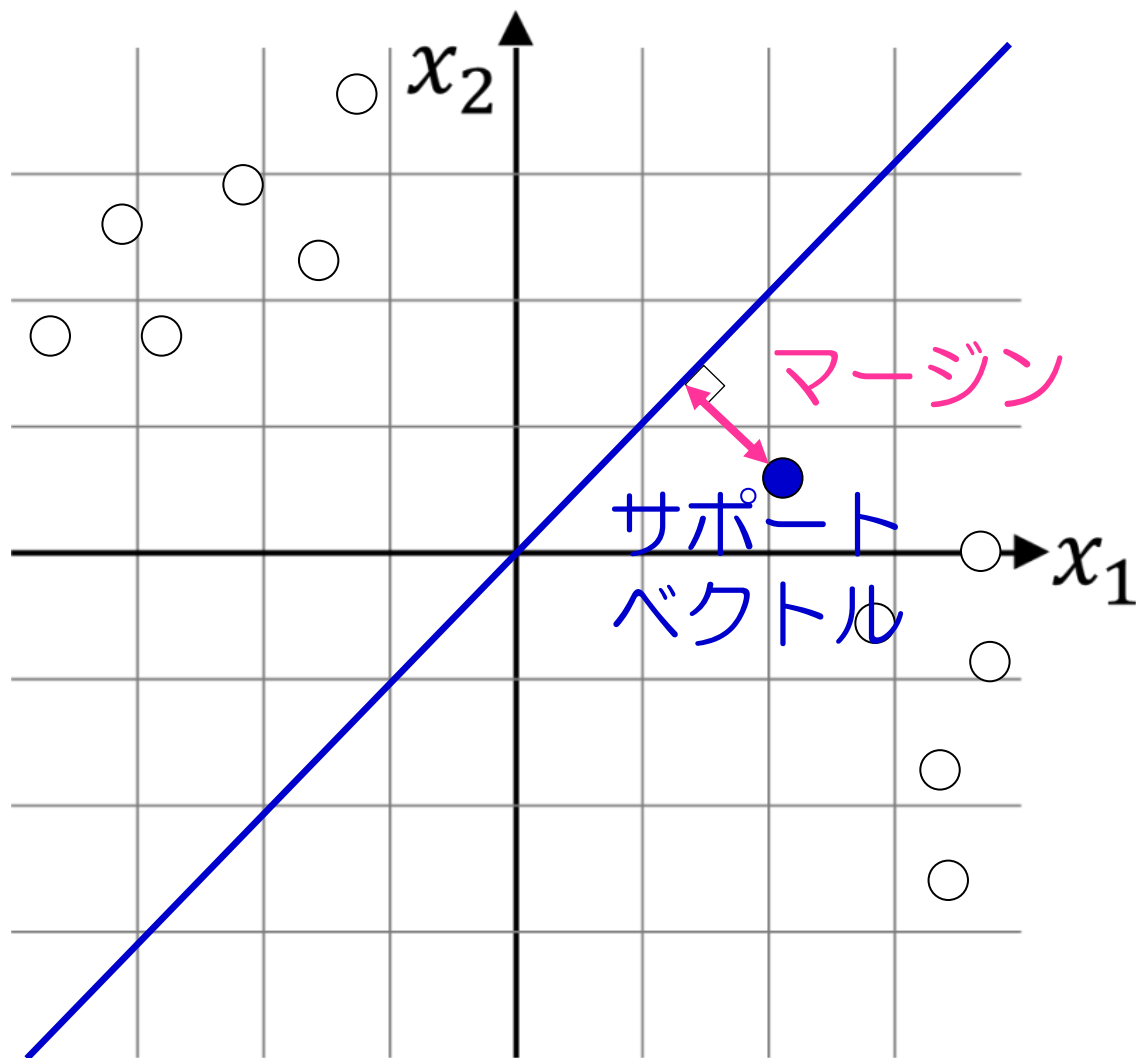
$$\{w | yw \cdot x_{sv} \geq \lambda\}$$

x_{sv} : w が作る分離平面に最も近い事例
(正負の区別なし)

↑
サポートベクトル

この集合の中から、
一番いい重みベクトルを選びたい！

サポートベクトルとマージン



マージン

- マージン：

- 分離平面から，分離平面にいちばん近い事例（サポートベクトル）までの距離

- さっきの w の集合からマージンを一番大きくするような分離平面を選ぼう！



マージン最大化

マージン最大化(1/2)

$$y\mathbf{w} \cdot \mathbf{x}_{sv} \geq \lambda$$



$$y\mathbf{w} \cdot \mathbf{x}_{sv} - \mathbf{y}\mathbf{w} \cdot \mathbf{x}^* \stackrel{0}{=} \geq \lambda$$



$$y\mathbf{w} \cdot (\mathbf{x}_{sv} - \mathbf{x}^*) \geq \lambda$$



$$y|\mathbf{w}||\mathbf{x}_{sv} - \mathbf{x}^*|\cos\theta \geq \lambda$$

同符合



$\theta = 0^\circ \text{ or } 180^\circ$

$$|\mathbf{w}||\mathbf{x}_{sv} - \mathbf{x}^*| \geq \lambda$$

マージン最大化(2/2)

$$|w| |x_{sv} - x^*| \geq \lambda$$



$$|x_{sv} - x^*| \geq \frac{\lambda}{|w|}$$

← 定数

マージン

マージン最大化の意味

$$|x_{sv} - x^*| \geq \frac{\lambda}{|w|}$$

L2ノルムが小さい重みベクトルを選べばマージンが大きくなる

マージン最大化の意味

$$|x_{sv} - x^*| \geq \frac{\lambda}{|w|}$$

L2ノルムが小さい重みベクトルを選べばマージンが大きくなる

ある学生

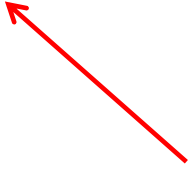


重みが小さくなると
マージンが広くなる
ってというのがピンと
こないです！

マージン最大化の厳密な意味

$$|x_{sv} - x^*| \geq \frac{\lambda}{|w|}$$

$\{w | yw \cdot x_{sv} \geq \lambda\}$ の中からL2ノルムが小さい重みベクトルを選べばマージンが大きくなる可能性がある



ベクトル w の長さが変わっても傾きが変わらなければ、分離平面は動かないので、マージンも変わらない

そうしたらまたこの漢が...

それは、 w の大きさがマージンの大きさと関係あることの説明になってないです！



証明しよう

まず、これ↓は自明

$$W_{min} = \{w | yw \cdot x_{sv} = \lambda\} \text{ としたとき,}$$
$$\{w | yw \cdot x_{sv} \geq \lambda\} = \{cw | w \in W_{min}, c \geq 1\}$$

自明な理由：

$yw \cdot x_{sv} \geq \lambda$ が成り立っているベクトルの長さを
どんどん短くしていけば

必ず、 $yw \cdot x_{sv} = \lambda$ となる w にたどり着く
このとき、分離平面の傾きは一切変わらない

つまり、 $|w|$ を小さくしてもマージンが変わらないことがある
でも、 W_{min} を求めることで、 w の候補数が絞り込める

$$W_{min} = \{\mathbf{w} | y\mathbf{w} \cdot \mathbf{x}_{sv} = \lambda\}$$

の中にある \mathbf{w} は全て異なる傾きを持っている
(傾きが同じなら, 別の \mathbf{w} の定数倍で表されるから)

そして, 同じ傾きを持つ \mathbf{w} の中で
大きさが最小のものだけが入っている

この中から最小のものを探すが
マージン最大化の本来の意図

$$y\mathbf{w} \cdot \mathbf{x}_{sv} = \lambda$$



$$|\mathbf{w}| |\mathbf{x}_{sv} - \mathbf{x}^*| = \lambda$$

トレードオフ

より, λ は定数であり,
 \mathbf{x}_{sv} および \mathbf{x}^* は
 \mathbf{w} の傾きで一意に決まっている

つまり, この状況でマージンを
 大きくしようと思った場合,

$$W_{min} = \{\mathbf{w} | y\mathbf{w} \cdot \mathbf{x}_{sv} = \lambda\}$$

の中から, $|\mathbf{w}|$ が最小となる \mathbf{w} を選ぶ必要がある!

ちなみに、 $|w|$ を最小化するのは、
L2正則化の効果もある。

途中で言った通り、 $|w|$ が無限に長くできるから、
できるだけ短い w を選ぶようにすることで、
候補を絞り込んでいる。

つまり、

$$y\mathbf{w} \cdot \mathbf{x}_{sv} \geq \lambda$$

の下での $|w|$ の最小化



- ベクトル w の長さについて、
正則化の効果
- ベクトル w の傾きについて、
マージン最大化の効果

を持っている

SVMの目的関数

$$\max(0, \lambda - y\mathbf{w} \cdot \mathbf{x}) + |\mathbf{w}|$$

を最小にする \mathbf{w} を探すのがSVMの学習

SVMまとめ

実はほとんど

パーセプトロン

ヒンジロス



マージン最大化!

重みは最小に...

宿題

- スコア関数を $w \cdot x$ のようにすると，分離平面が必ず原点に拘束されることを証明しなさい
- スコア関数に以下のようにバイアス項を設けると，原点に拘束されず，素性空間の任意の位置に分離平面が引けるようになることを説明しなさい

$$score(x) = w \cdot x - b$$