

機械学習入門

～教師あり学習について～

講師：高岸茉莉子

注目される人工知能（AI）技術

**データサイエンス：
国のAI戦略で「全大学生が習得」**

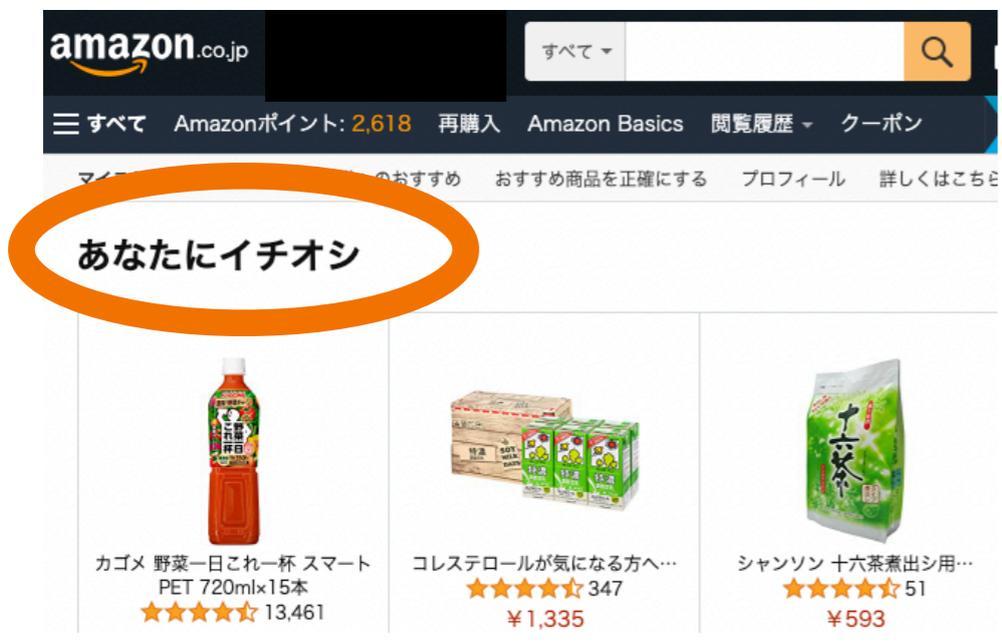


**デジタル人材* 「別枠採用」が3割
主要企業の来春新卒**

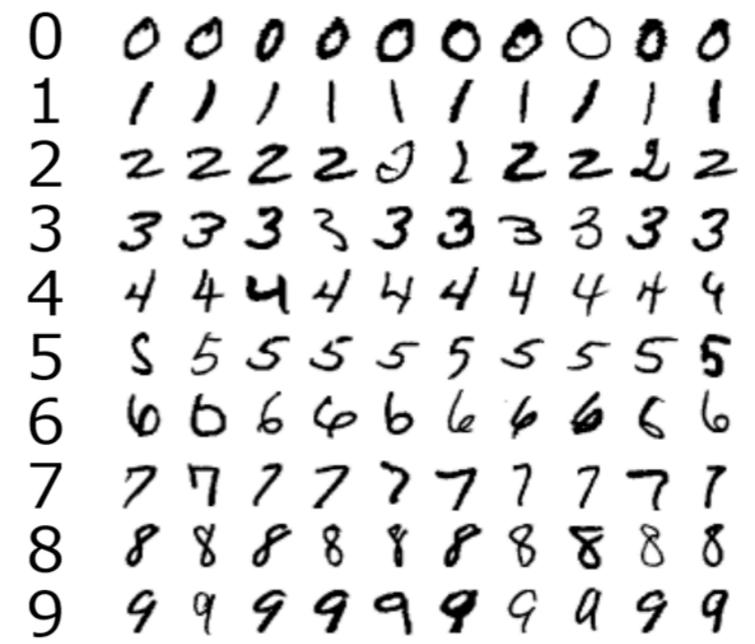
*（記事では）データ分析や人工知能（AI）
などの専門人材を指す

"AI" (機械学習) 技術の例

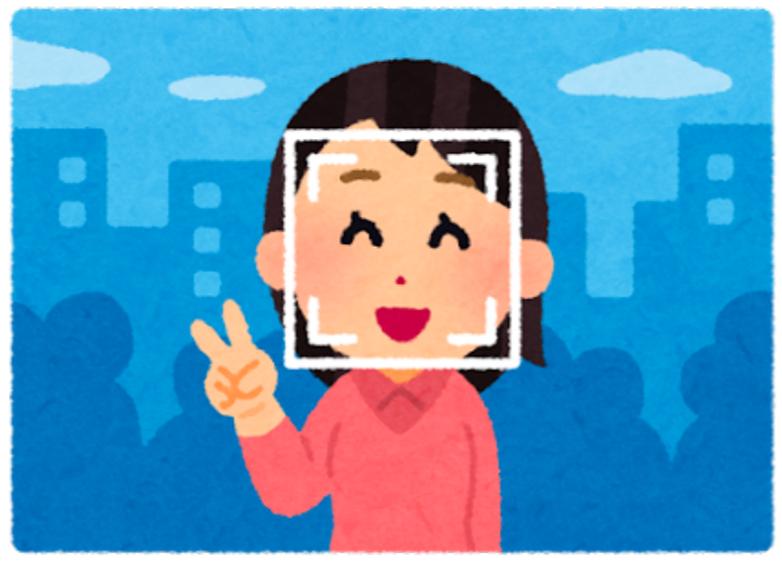
推薦システム



手書き文字認識



カメラの顔認識

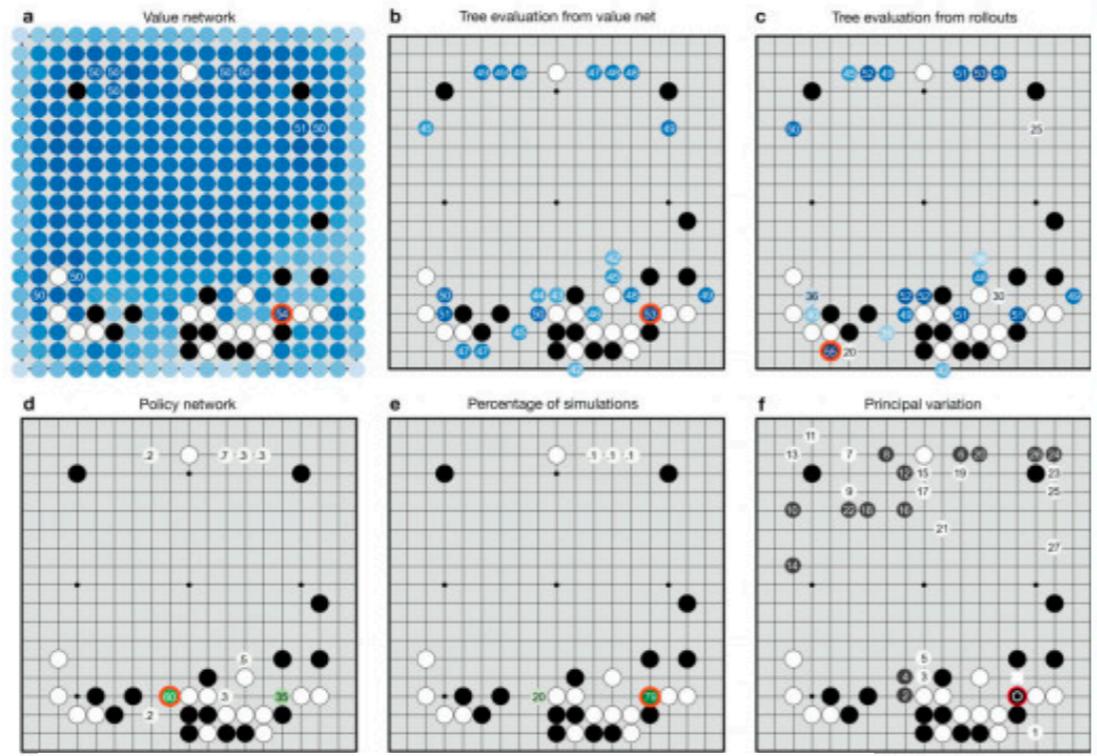


音声認識 (Siri)



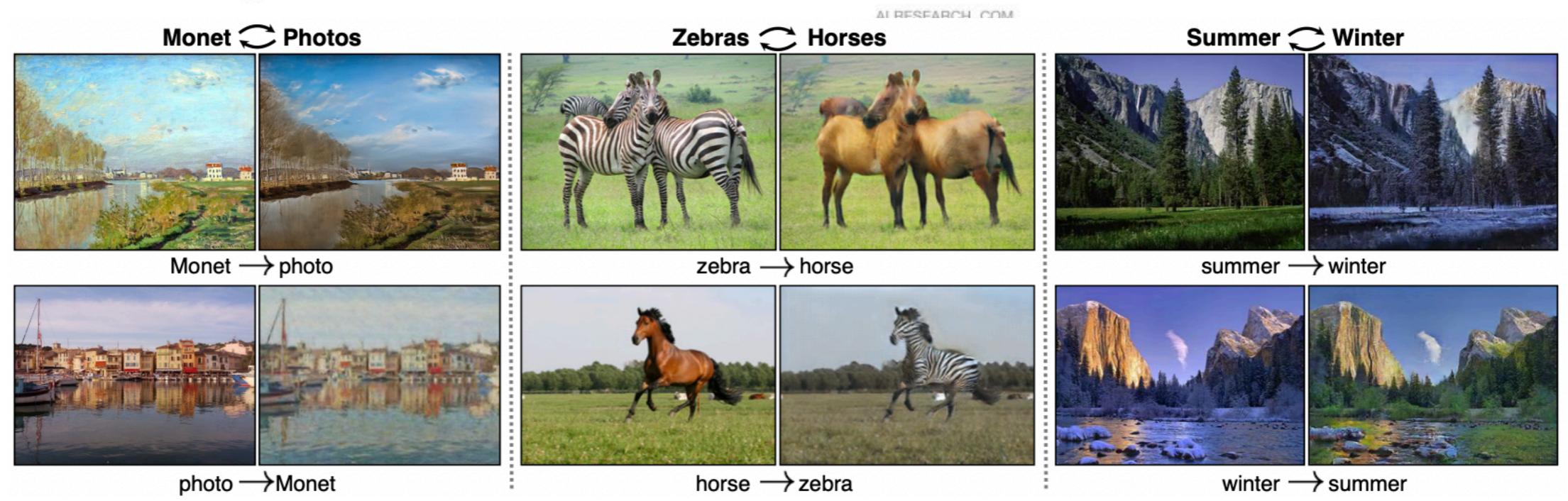
"AI" (機械学習) 技術の例

→プロ棋士を破った囲碁のAI (Alpha Go)



好みのイラストをAIで生成

↓ 画像の特徴を変換



機械学習とは

- 機械学習：多くの人工知能（AI）の基盤となっている技術。
 - 機械学習の技術を理解するには、数学がとても重要。
 - また基本的な機械学習の技術であれば、高校数学の範囲でも、そのメカニズムを直感的に理解することも可能
- この講義では、機械学習のカテゴリの1つ「教師あり学習」に注目し、その一部を紹介。

機械学習とは（教師あり学習の場合）

教師あり学習とは...予測や判別等のために使う関数を、得られたデータから"学習"すること。

- 中高で習った"関数"：入力 x に対し，何らかの加工して， y を出力（例： $f(x) = 2x + 3 = y$ ）

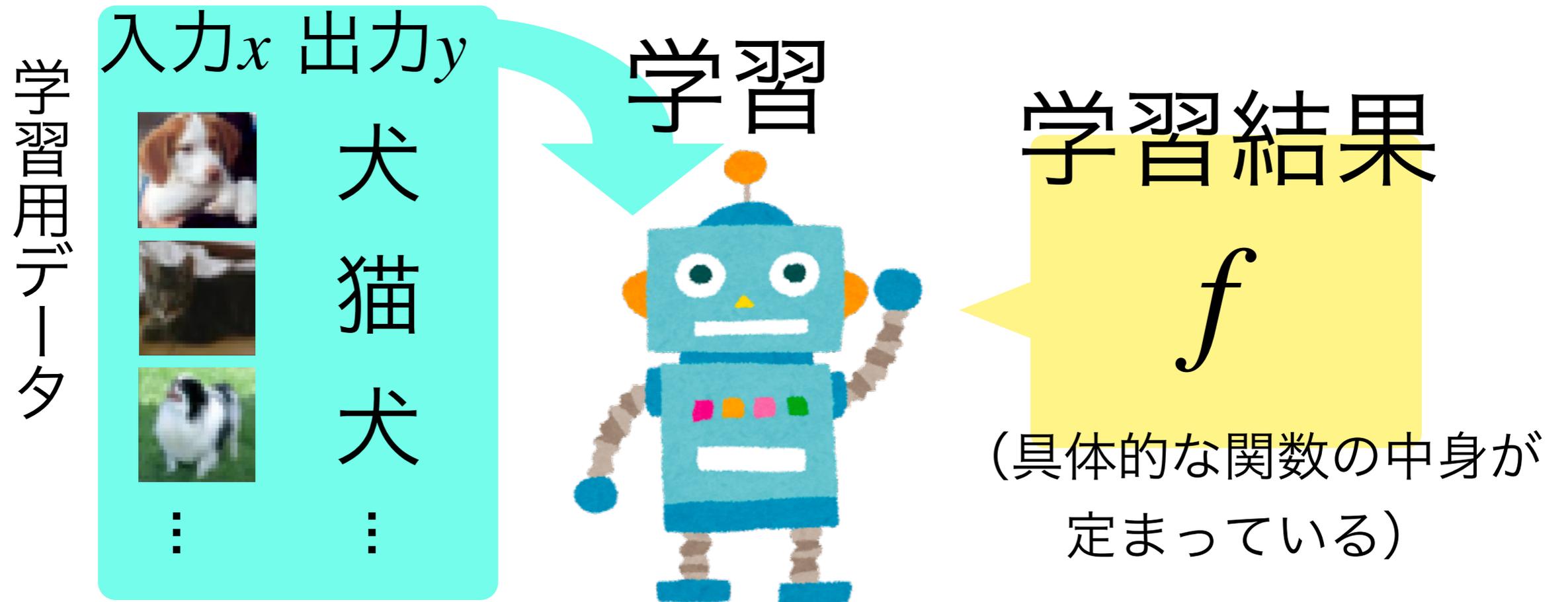


- 機械学習（教師あり学習）で扱う"関数"：入力から出力を予測。関数 f の中身は未知。

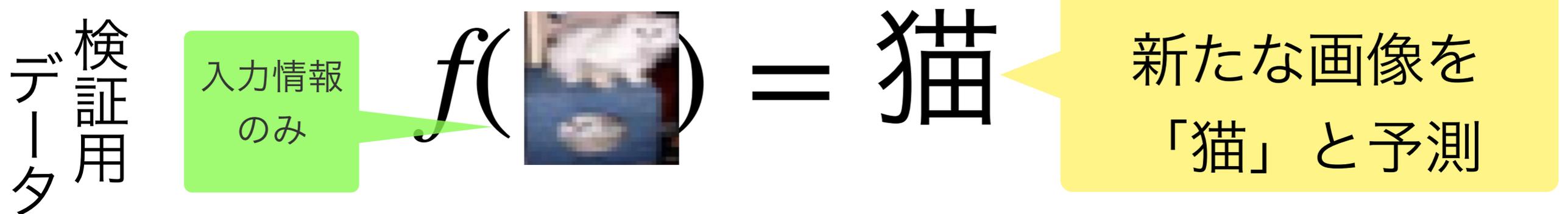


教師あり学習の流れ

①学習段階



②検証段階



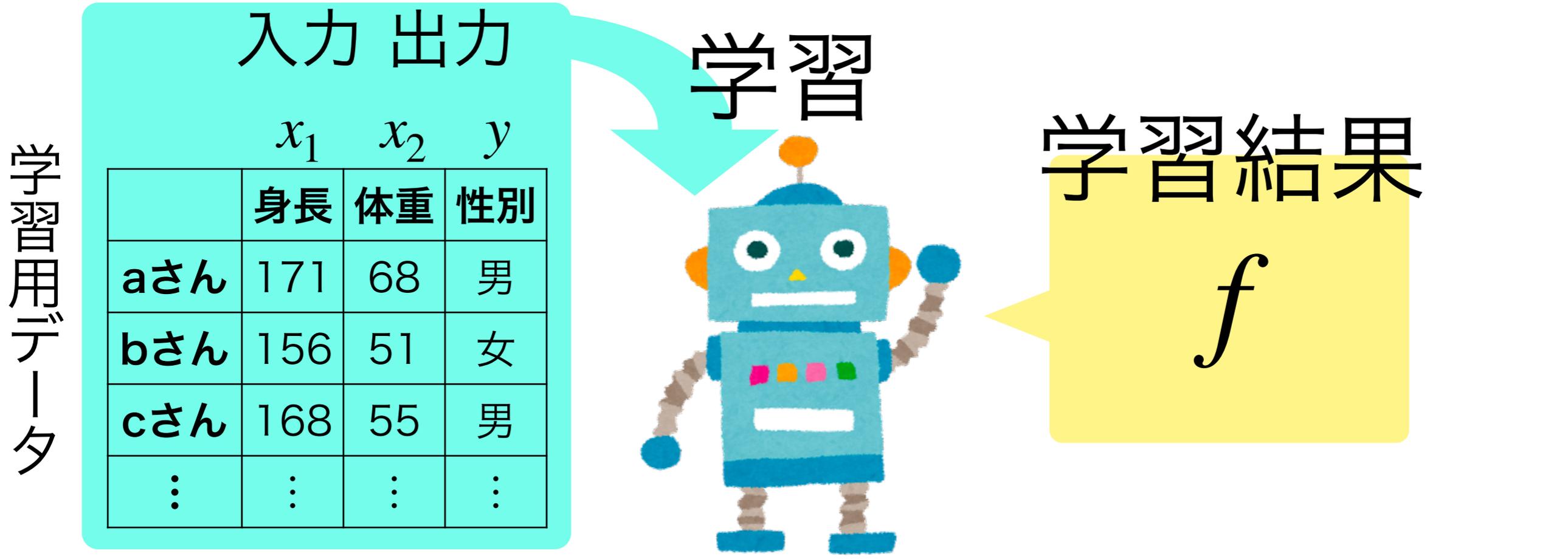
"関数の学習"とは？

～分類のための

教師あり学習の場合～

具体例：身長・体重から性別予測

- 目的：身長・体重から性別を予測する関数 f を学習。



- 学習結果を用いると...

検証用データ

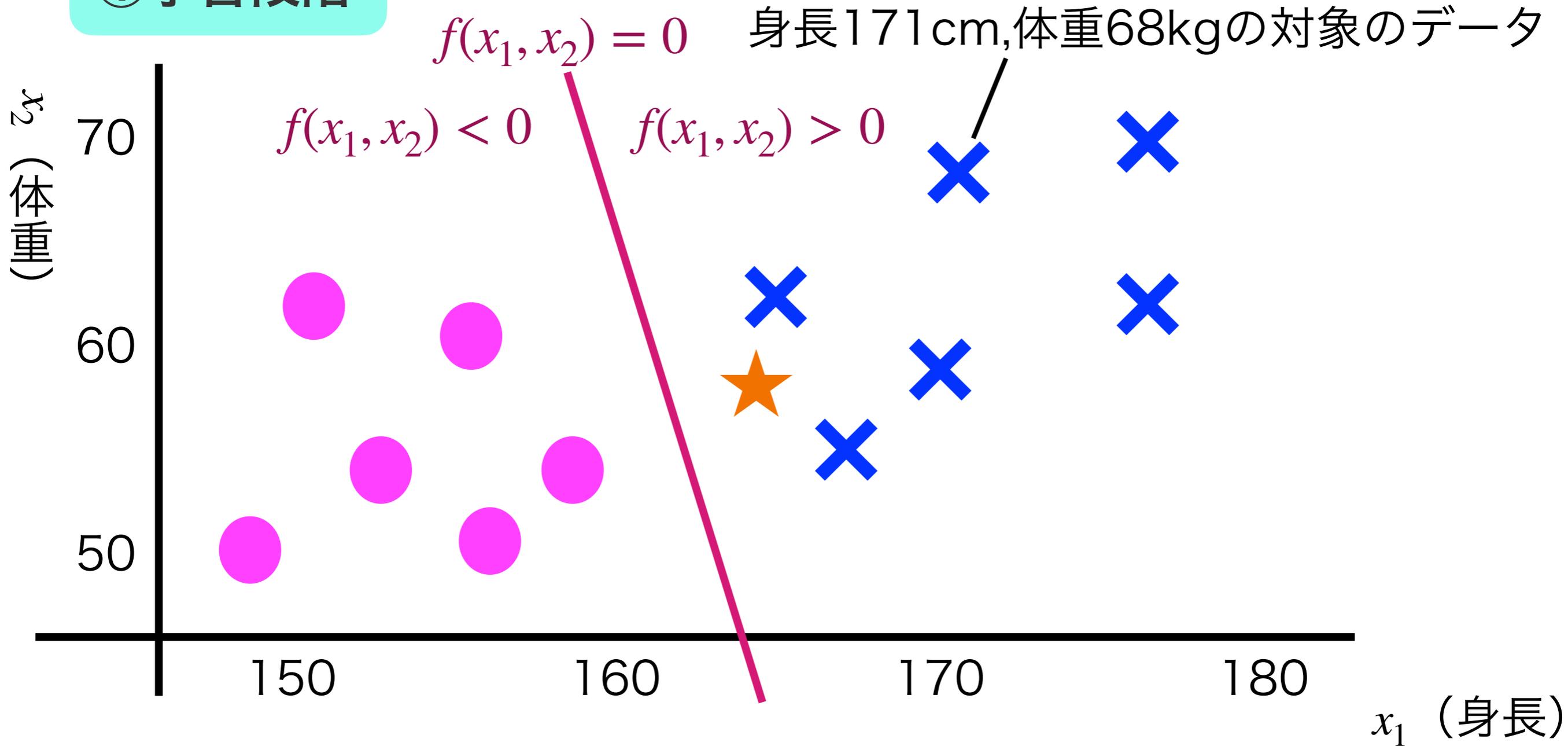
入力情報のみ

$$f(165, 60) = \text{男}$$

身長・体重を入力すると性別を予測

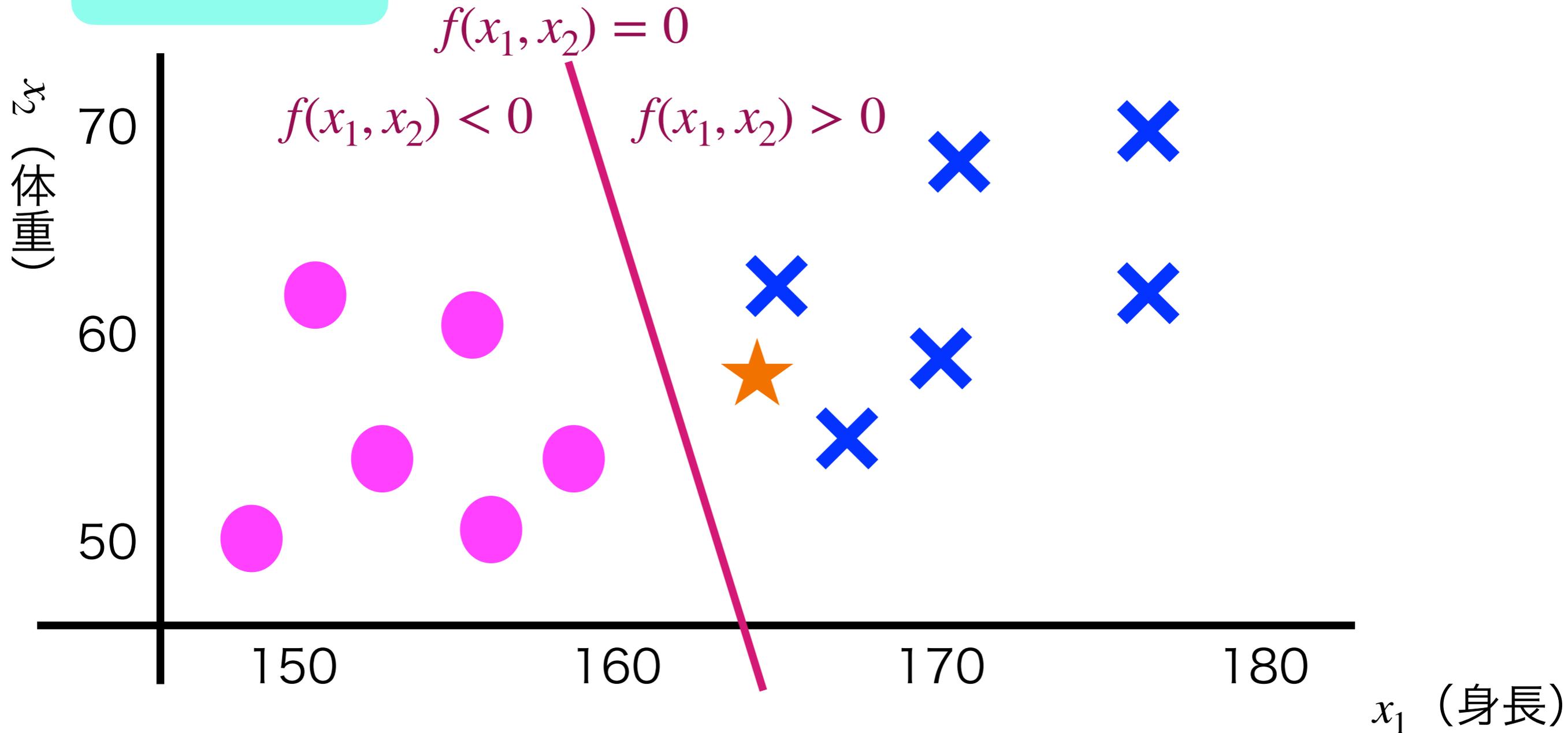
「関数 f を学習する」とは？

①学習段階



「関数 f を学習する」とは？

②検証段階



zさん（性別の情報なし）は身長164，体重58.

$f(164, 58) = 0.01 > 0 \rightarrow$ zさんは男性と予測

「関数 f を学習する」とは？

- Q,境界線の役割を果たす関数 f はどうやって得られるの？
- A, f を下記のように定義すればよい.

$$f(x_1, x_2) = w_0 + w_1x_1 + w_2x_2 \quad \dots\dots\dots(\star)$$

- ただし, w_0, w_1, w_2 は未知 (データから学習する) .
- 学習=具体的な値を得ること

- この場合, 境界線は下記を満たす x_1, x_2 の集合.

$$f(x_1, x_2) = w_0 + w_1x_1 + w_2x_2 = 0$$

→直線の方程式！

「関数 f を学習する」とは？

- Q, f が直線を意味することは分かったが, 「 f を学習する」とは結局どういうこと?
- 関数 f において, w_0, w_1, w_2 は未知.

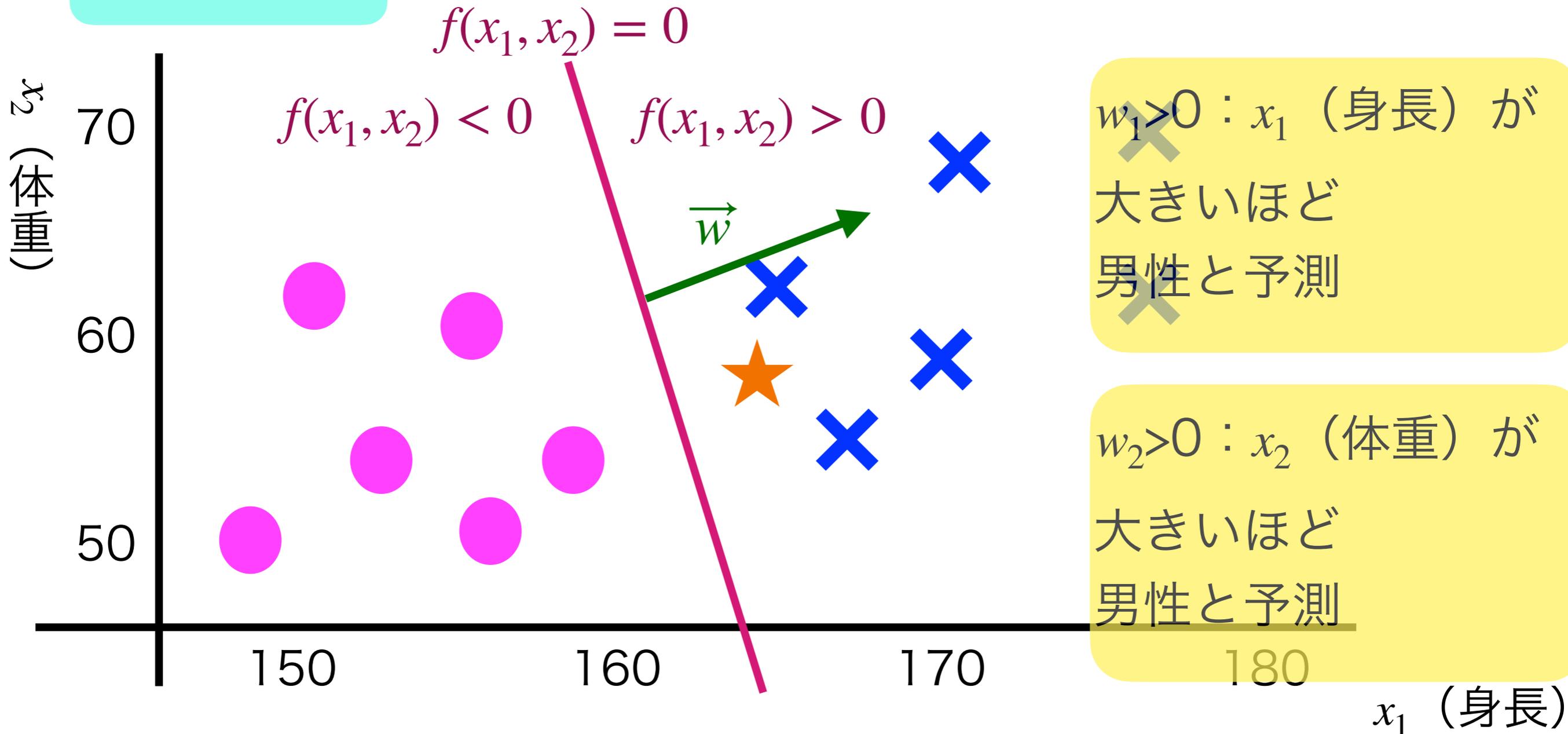
$$f(x_1, x_2) = w_0 + w_1x_1 + w_2x_2 = 0$$

- ここで $\vec{w} = (w_1, w_2)$ は直線に垂直なベクトル (法線ベクトル)
- つまり, \vec{w} を学習 (=ベクトルを具体的に定める)
→ \vec{w} と垂直になる直線が定まる (これが境界線に！)

境界線が, うまい具合に男女予測できる場所に引かれるよう
 \vec{w} をうまく定める (=学習する) のが, よい機械学習の方法！

「関数 f を学習する」とは？

②検証段階



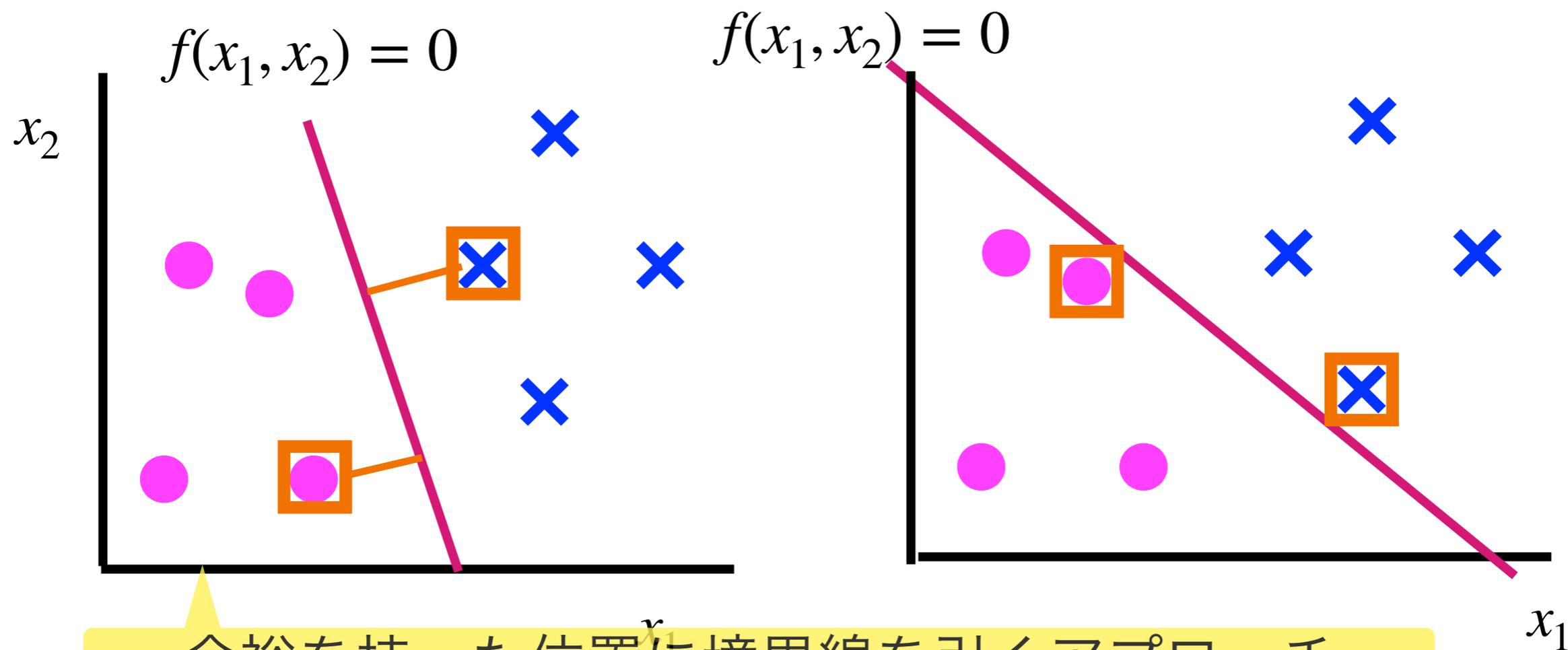
zさん (性別の情報なし) は身長164, 体重58.

$f(164, 58) = 0.01 > 0 \rightarrow$ zさんは男性と予測

様々な境界線の決め方

サポートベクターマシン

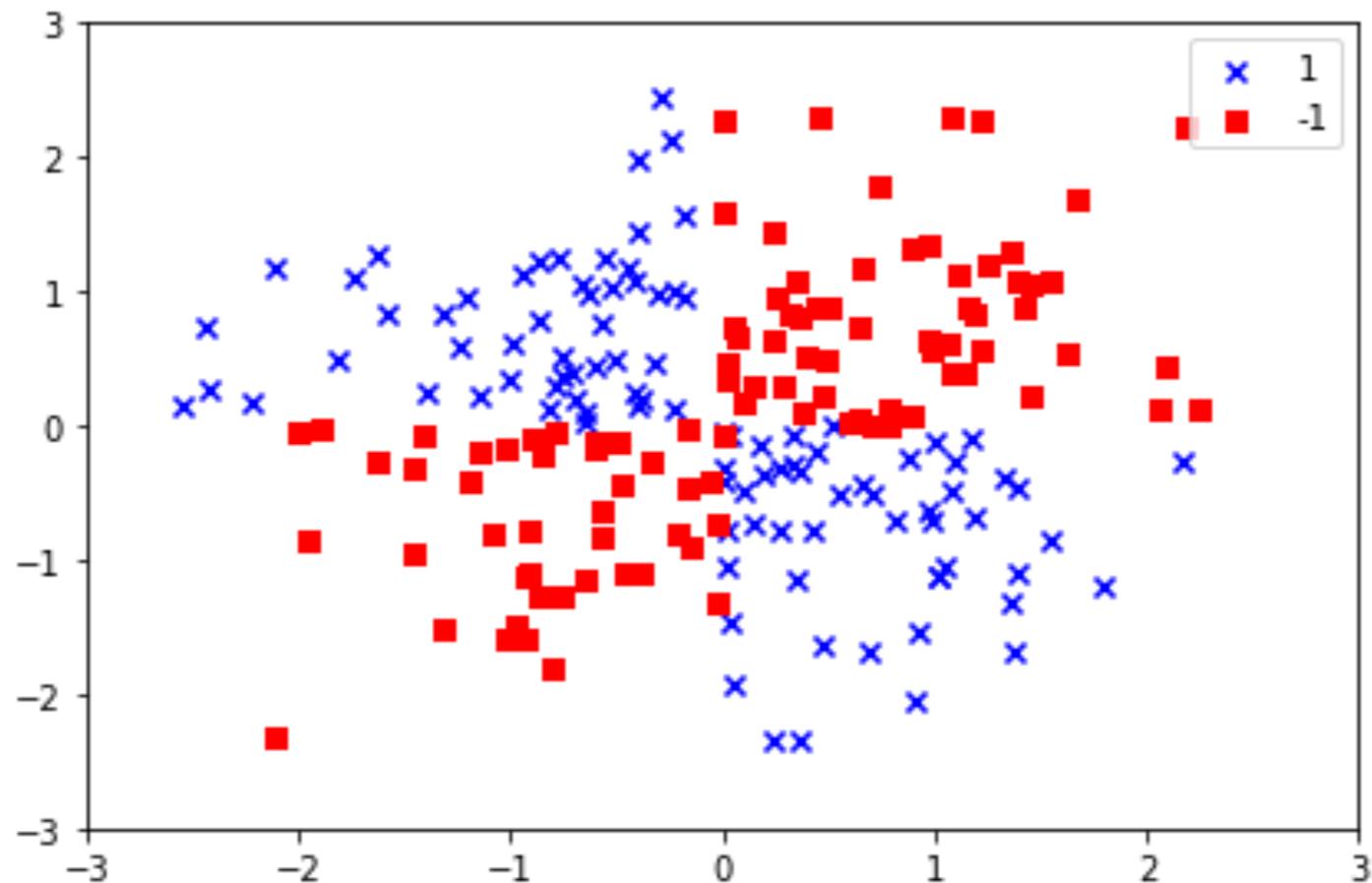
- 関数 f （境界線）の推定の方法は様々.
- Q, 新たなデータが得られた時, 分類の失敗が少なそうな境界線はどっち?



余裕を持った位置に境界線を引くアプローチ
→サポートベクターマシン (SVM)

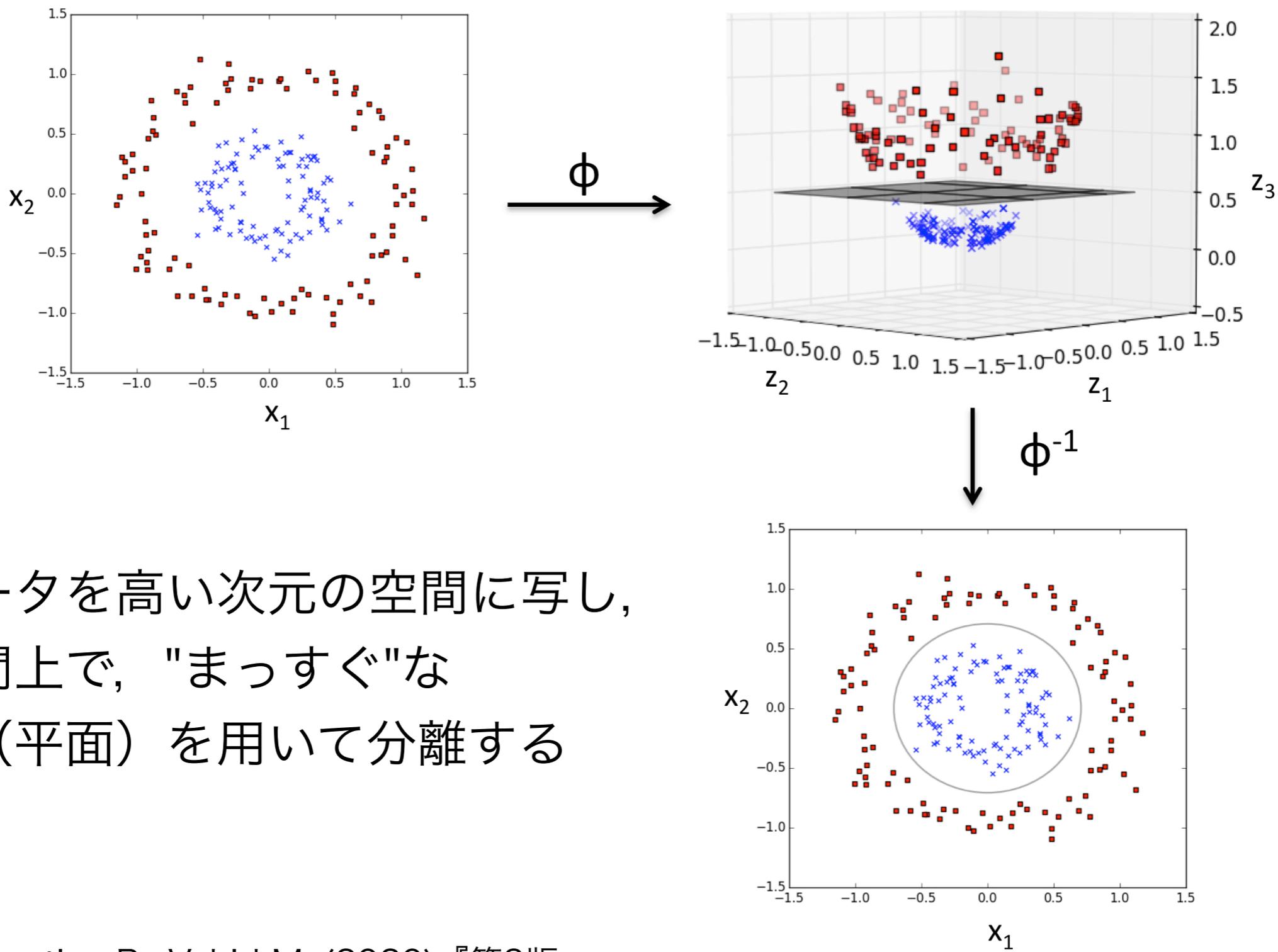
さらに複雑な場合

- 下記のような場合：直線で分類は難しい



- →カーネル法と呼ばれるトリックを使えば、「直線」での分類が可能に.

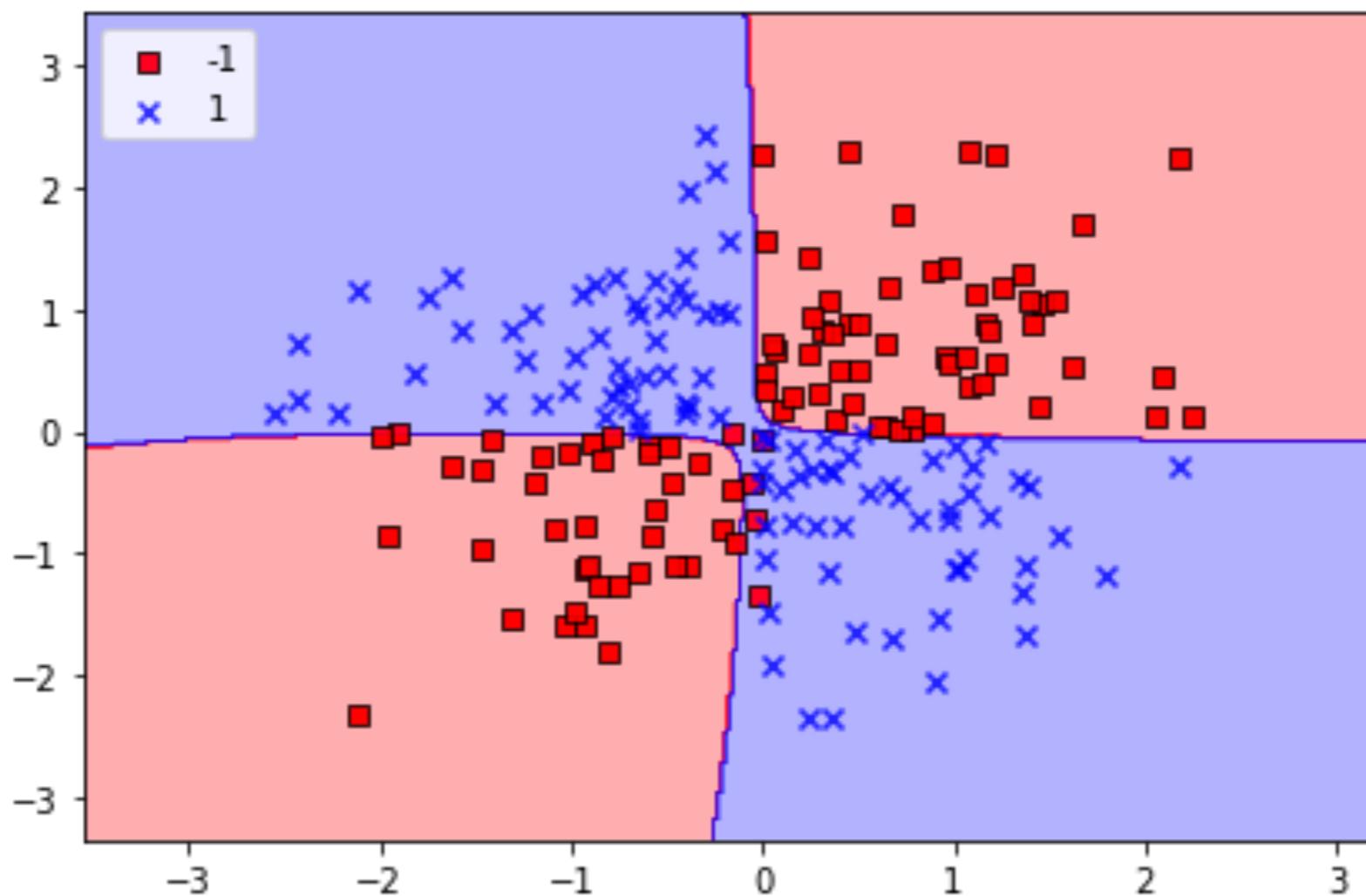
カーネル法のイメージ



入力データを高い次元の空間に写し、
その空間上で、"まっすぐ"な
境界線（平面）を用いて分離する

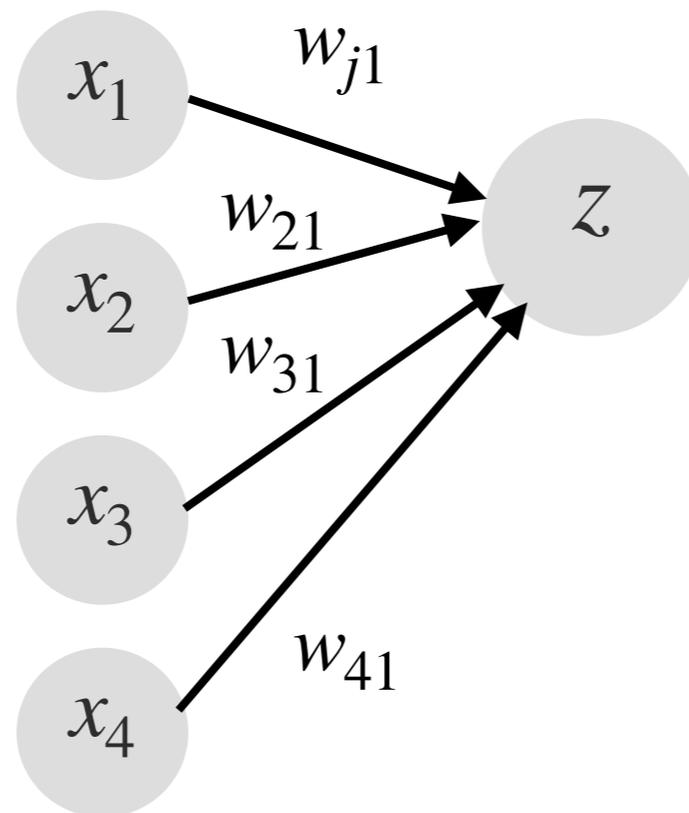
SVM with カーネル法の適用結果

- うまく分類できている。



より複雑な分類に向けて
～深層学習～

深層学習の構造（一部）



先ほどの式
 $(f(x_1, x_2) = w_0 + w_1x_1 + w_2x_2)$ を
 少し複雑にしたもの

$$z = h(w_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4), \quad (j = 1, 2, 3)$$

- h : 活性化関数（ある閾値を超えると，神経細胞（脳内にある細胞）が信号を発することをモデル化した関数）

まとめ

- 分類のための「教師あり学習」の手法の一部を紹介した。
- どの手法も、データを分類するための「うまい境界線を引く」ことを目的としている
 - 「うまい」の意味が手法ごとに異なる。
- 結局、分類のための教師あり機械学習とは：
「うまい境界線を引く」ことを、様々な数学的手段を用いて達成すること。
 - やりたいこと自体は、小学生でもイメージできる単純なこと。
 - ただし数学的に達成することが難しい

まとめ

- 分類のための「教師あり学習」の手法の一部を紹介した。
- どの手法も、データを分類するための「うまい境界線を引く」ことを目的としている。

岡山大学工学部数理データサイエンスコースで、
そのメカニズムを理解しつつ、

- 様々な機械学習法を使いこなせるようになるよう「うまい境界線を引く」ことを、様々な数式的手段を用いて達成すること。
 - やりたいこと自体は、小学生でもイメージできる単純なこと。
 - ただし数学的に達成することが難しい

参考文献

- Alpha Go, <https://www.deepmind.com/research/highlighted-research/alphago> (2022/7/10閲覧)
- crypko, <https://crypko.ai/> (2022/7/10閲覧)
- Sebastian R., Vahid M. (2020) 『第3版 Python機械学習プログラミング 達人データサイエンティストによる理論と実践』 インプレス
- Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).
- 平井 有三 (2012) 『はじめてのパターン認識』 森北出版
- 八谷大岳 (2020) 『機械学習スタートアップシリーズ ゼロからつくるPython機械学習プログラミング入門』 講談社