# B3セミナー

M1　新井　棟大

# 目次

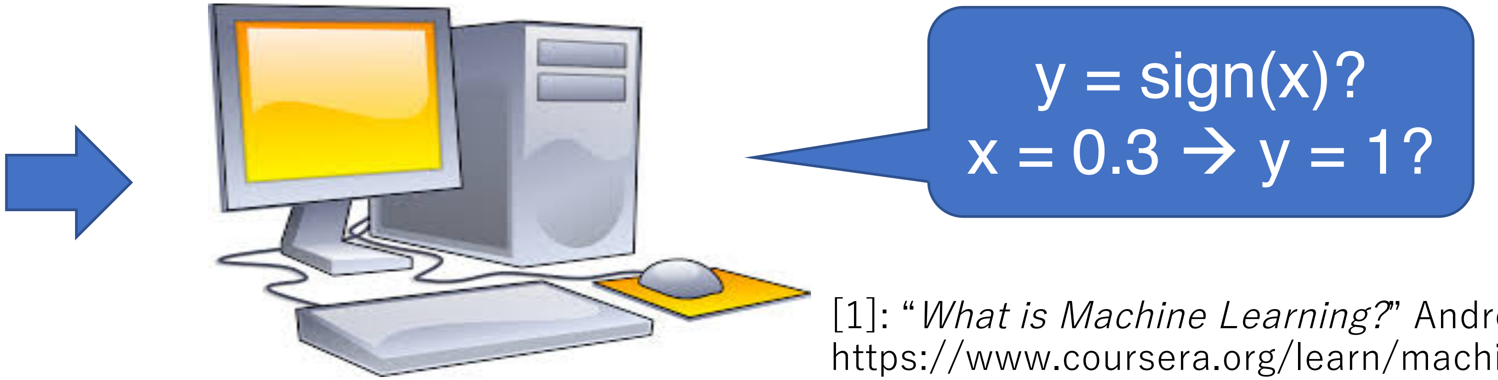
- 機械学習とは？
- 線形分類器
  - 定式化・損失関数
  - 勾配降下法
  - 限界

# What is machine learning (機械学習)?

- “*Machine learning is the science of getting computers to act without being explicitly programmed.*” [1]
  - 知識/パターンを勝手に発見するプログラムの開発

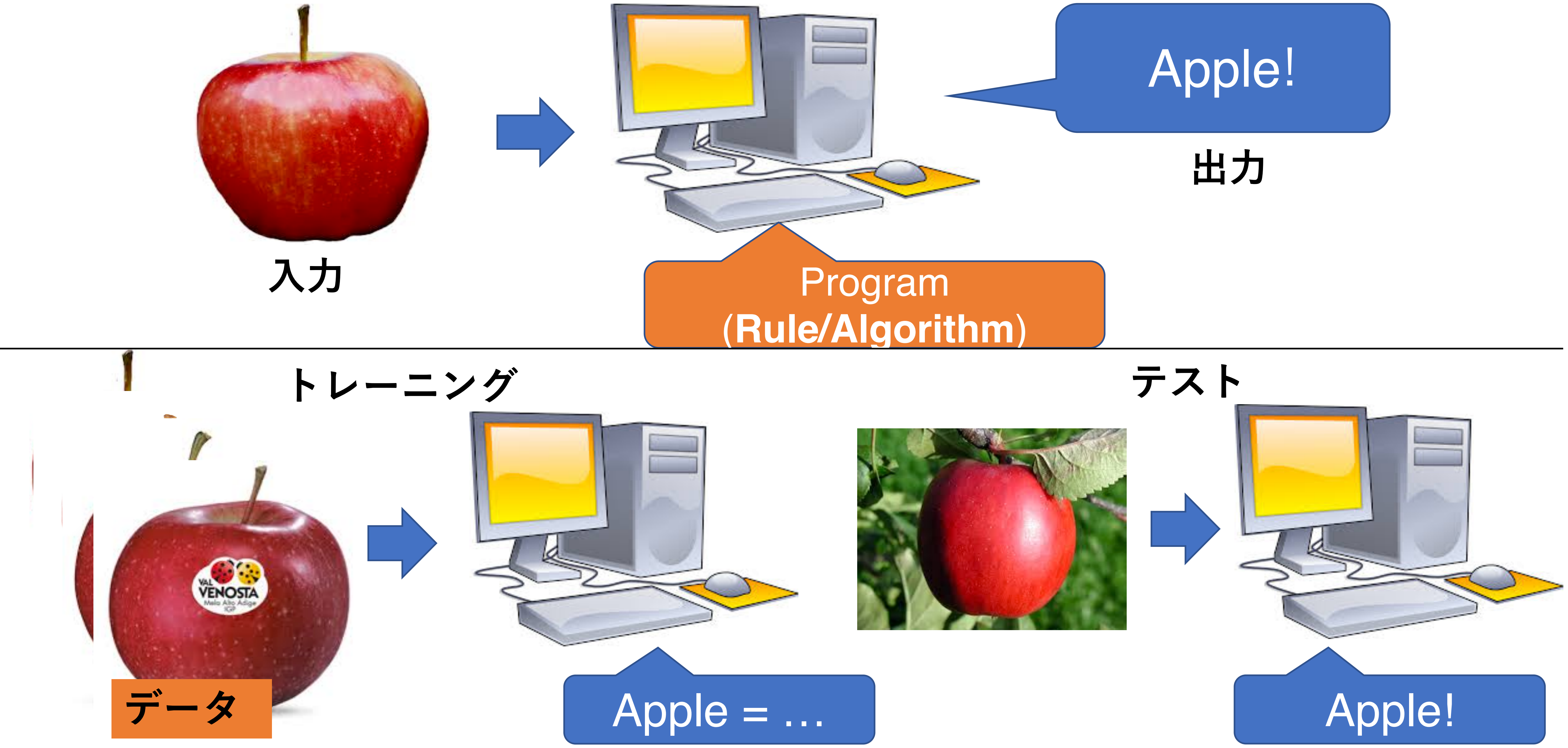
- (教師あり)学習
  - データ$\{x_1, x_2, \dots x_N\}$, ラベル$\{y_1, y_2, \dots y_N\}$ → y = $f(x)$なる$f$を求める
  - e.g.) x = {-2, -1, 2, 3}, y = {-1, -1, 1, 1} → f = sign(x)

| x | y |
| --- | --- |
| -2 | -1 |
| -1 | -1 |
| 2 | 1 |
| 3 | 1 |




[1]: “*What is Machine Learning?*” Andrew Ng. https://www.coursera.org/learn/machine-learning/lecture/Ujm7v/what-is-machine-learning

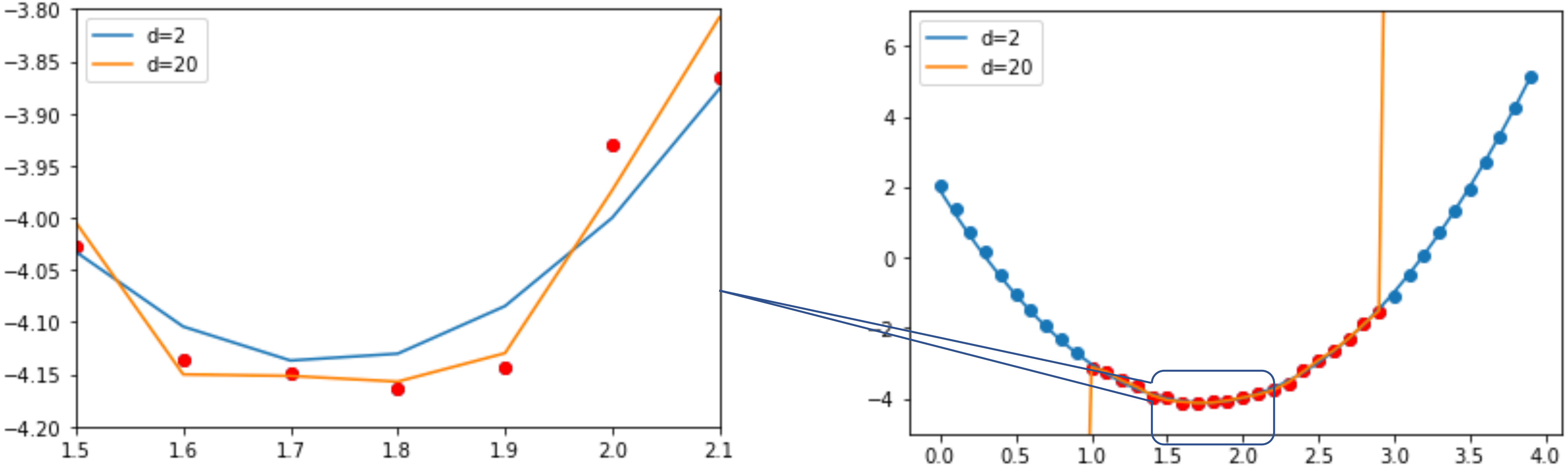
# 機械学習と普通のプログラムの違い



# 機械学習の難しい点: 過学習

- 全部覚えるのではなく、妥当な「傾向」を学習したい
  - データにはノイズが乗っている(と仮定する)ことが多い
  - 得られるデータ(サンプル)は母集団の一部
  - 知りたいのは未知データのラベル

- **過学習**
  - トレーニングデータに適応しすぎた為に未知データに対するラベルの予測精度が落ちてしまう状態



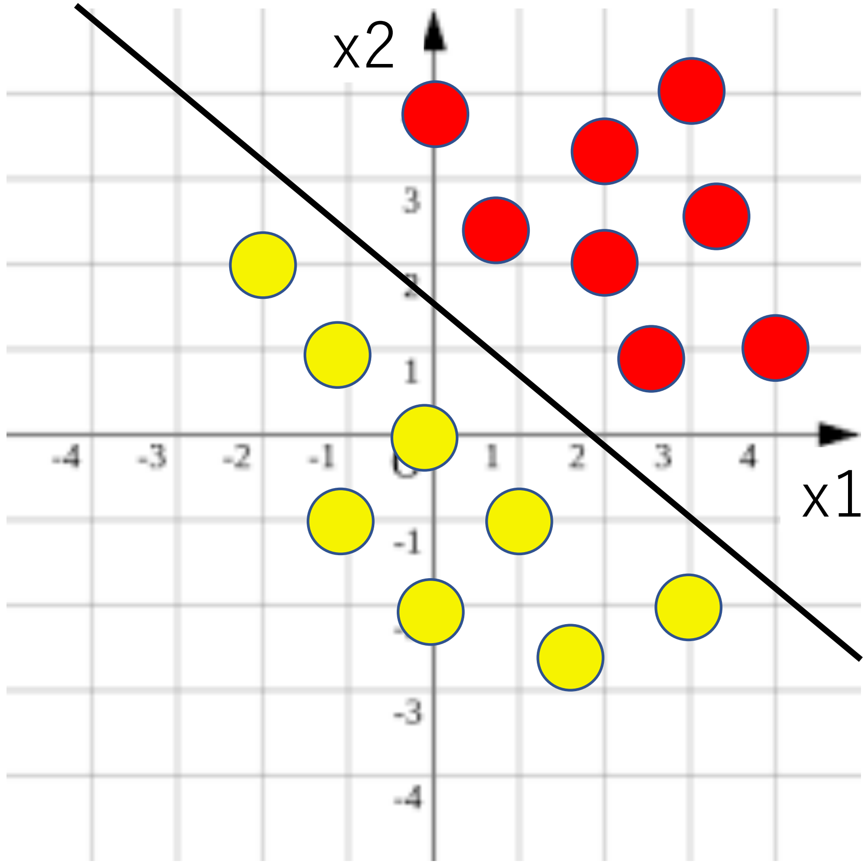
# 線形識別器

- Apple or Banana?
  - 2クラス分類問題
  - 画像からいきなりでは難しいので**特徴量**を計算して分類
  - データ: **x** = [x1, x2], x1 = 色, x2 = 形
  - ラベル: y = 1 (Apple) or -1 (Banana)

- 線形識別器
  - 平面を直線で分割してサンプルがどの領域に割り当てられているかにより分類
  - $f(x) = sign(\boldsymbol{w}^T \boldsymbol{x})$
  - SVMはこの1種

[-1,-1]

[2,2]




* $\boldsymbol{x} = [x_1, x_2, 1]$

# 線形識別器: 定式化

- パラメーター$\mathbf{w}$の求め方: 識別ロスの最小化
  - 識別の間違えの大きさを定める**損失関数**を定義

$$\mathrm{E}(\boldsymbol{w}) = \sum_i \frac{1}{2}(1 - y_{\mathrm{i}}\mathrm{sign}(\boldsymbol{w}^T\boldsymbol{x}_{\mathrm{i}}))$$

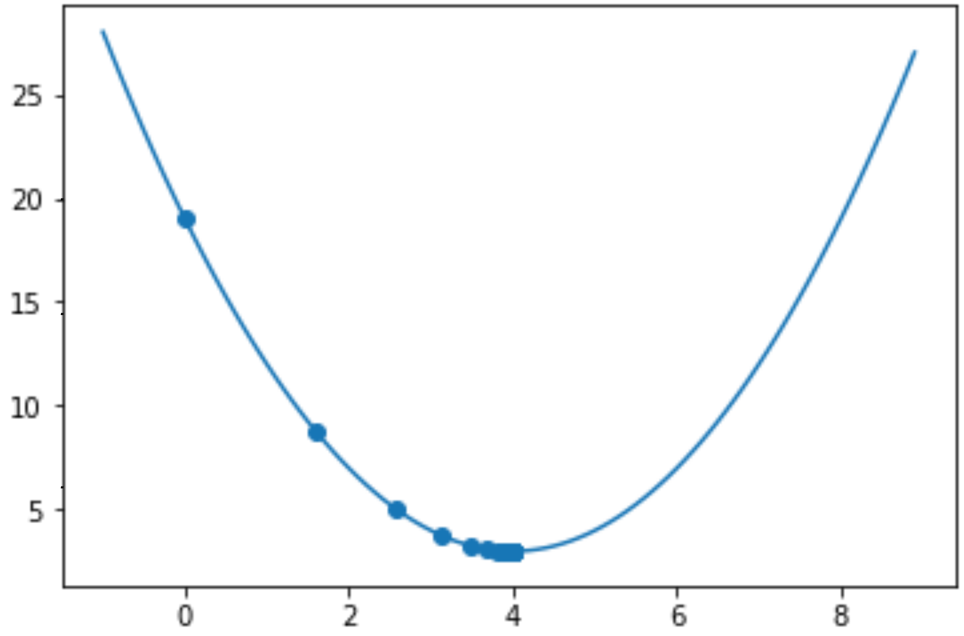
※Σ内は $f(\boldsymbol{x}_{\mathrm{i}}) = y_i$ なら0, 違えば1

  - これでは微分ができないので通常は以下のSigmoid lossなどの代理損失を用いる

$$\mathrm{E}(\boldsymbol{w}) = \sum_i \log\{1 + \exp(-y_i(\boldsymbol{w}^T\boldsymbol{x}_{\mathrm{i}}))\}$$

# 勾配降下法

- 関数の最適化手法の1つ。$\frac{\mathrm{df}}{\mathrm{d}\boldsymbol{x}}(\boldsymbol{x})$が$\boldsymbol{x}$ におけるfの最急降下方向であることを使った手法。
- アルゴリズム
  - 適当な初期位置$\boldsymbol{x_0}$を決める。
  - 以下を$x_{i+1} \approx x_i$ となるまで繰り返す。
    - $f$の$x_i$における微分$\frac{\mathrm{df}}{\mathrm{d}\boldsymbol{x}}(\boldsymbol{x_i})$を求める。
    - $x_{i+1} \leftarrow x_i - \alpha \frac{\mathrm{df}}{\mathrm{d}\boldsymbol{x}}(\boldsymbol{x_i})$としてxを更新する。
- 例
  - $f(x) = (x - 4)^2 + 3, \alpha = 0.2$



# 線形識別器: 更新式

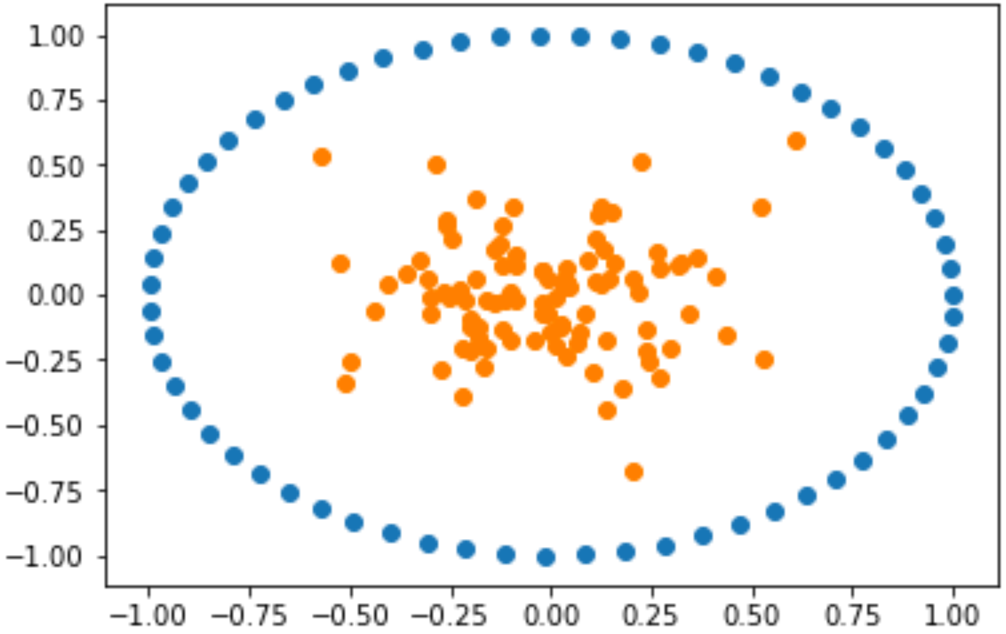
- 線形識別器の$\boldsymbol{w}$は損失関数を最小化するように定める。
  - $\mathrm{E}(\boldsymbol{w}) = \sum_i \log\{1 + \exp(-y_i(\boldsymbol{w}^T \boldsymbol{x}_{\mathrm{i}}))\}$
  - $\frac{\mathrm{d}}{d\boldsymbol{w}}\mathrm{E}(\boldsymbol{w}) = \sum_i -\frac{\exp(-y_i(\boldsymbol{w}^T \boldsymbol{x}_{\mathrm{i}}))}{1+\exp(-y_i(\boldsymbol{w}^T \boldsymbol{x}_{\mathrm{i}}))} y_i \boldsymbol{x}_{\mathrm{i}}$

  ➢ $\boldsymbol{w} \leftarrow \boldsymbol{w} + \alpha \sum_i \frac{\exp(-y_i(\boldsymbol{w}^T \boldsymbol{x}_{\mathrm{i}}))}{1+\exp(-y_i(\boldsymbol{w}^T \boldsymbol{x}_{\mathrm{i}}))} y_i \boldsymbol{x}_{\mathrm{i}}$

- より複雑な識別器では計算量の問題/局所解の回避という観点から、**確率的勾配降下法**(SGD)を用いることも多い
  - $\boldsymbol{w} \leftarrow \boldsymbol{w} + \alpha \frac{\exp(-y_i(\boldsymbol{w}^T \boldsymbol{x}_{\mathrm{i}}))}{1+\exp(-y_i(\boldsymbol{w}^T \boldsymbol{x}_{\mathrm{i}}))} y_i \boldsymbol{x}_{\mathrm{i}}$　を各iについて繰り返す

# 線形識別器: 限界

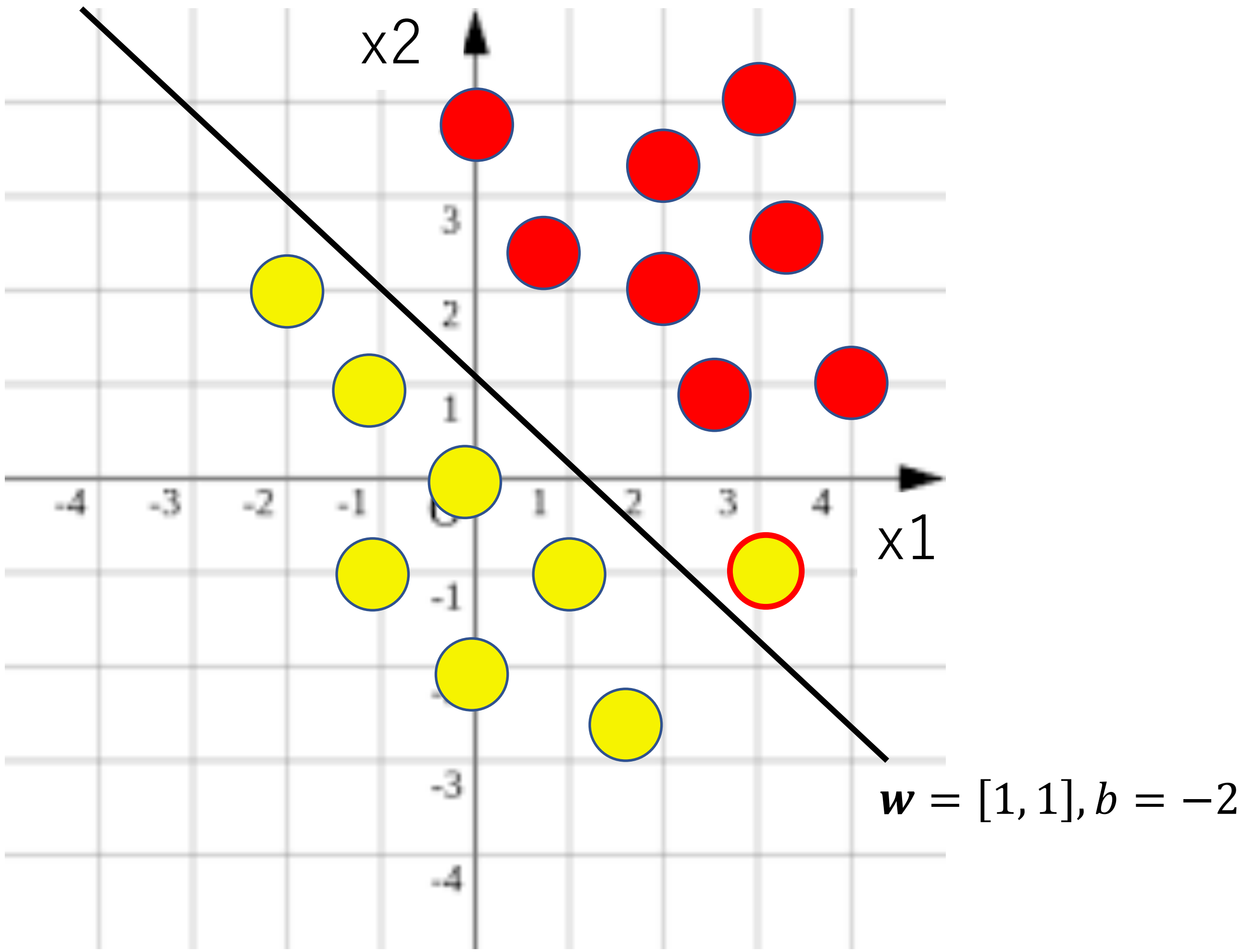
- 線形識別器は境界が線形でない分類は行えない
    - あらかじめデータを加工する/別の空間に射影することも可能ではある




- 非線形な境界面も表現できる識別器
    - ニューラルネットワークなど

# 前半のまとめ

- (教師あり)機械学習とはデータとラベルの組から、データに対するラベルを出力するような関数を見つける操作を指す。
- 線形識別器はクラス分類問題を解く方法(識別器)の1つである。
- 線形識別器は損失関数を勾配降下法を用いて最小化することでパラメータを定めることが出来る。
- 線形識別器では非線形な境界をもつクラスの分類は行えない。


x2
x1
-4
-3
-2
-1
1
2
3
4
3
2
1
-1
-3
-4
$\boldsymbol{w} = [1, 1], b = -2$