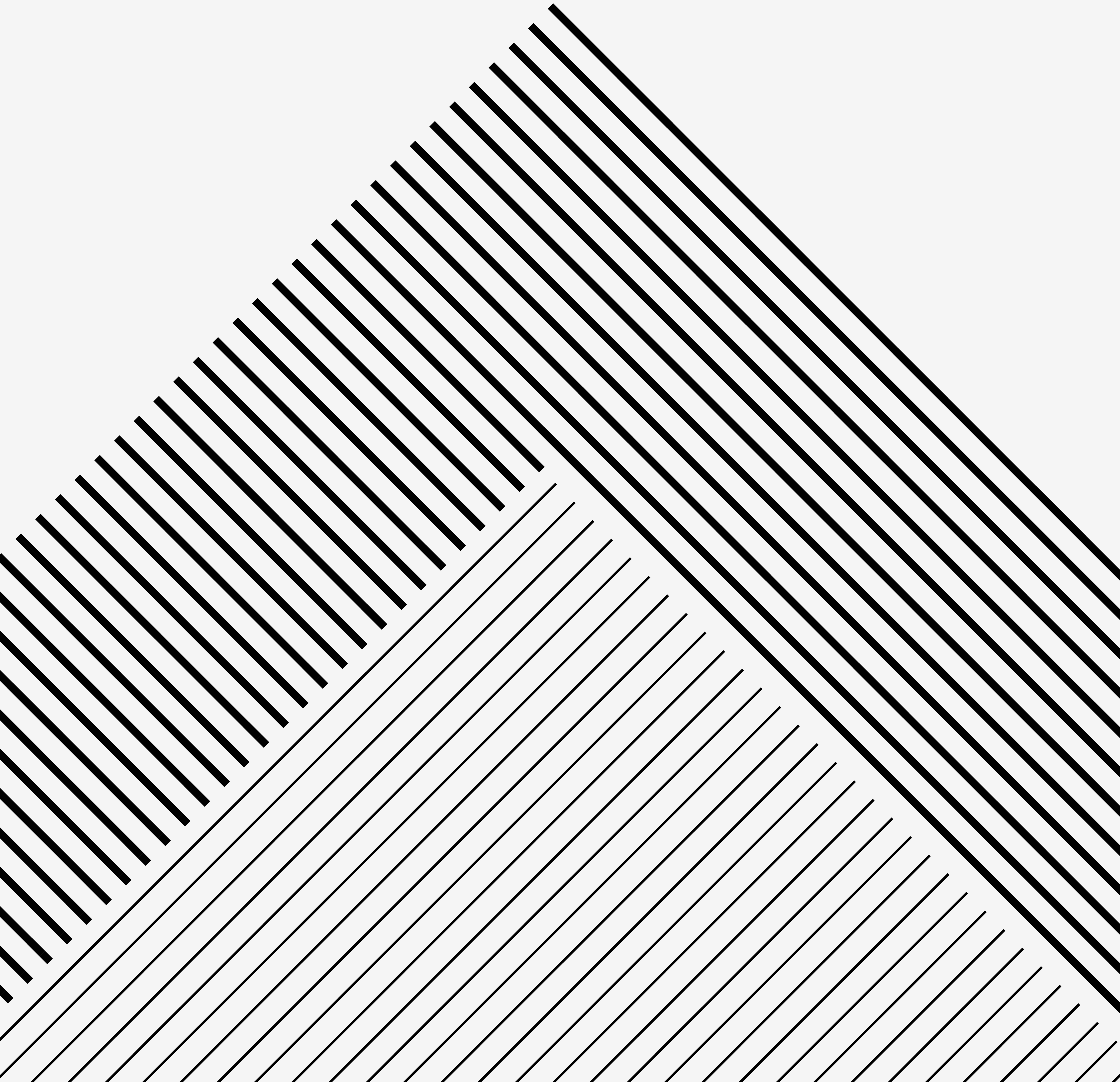


week 3

week 2までは
回帰問題

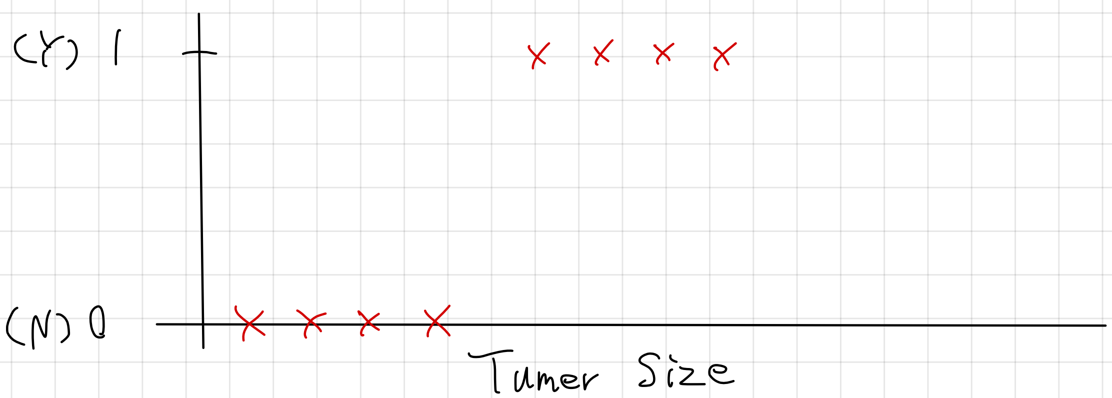
- 分類問題

  - Emailがスパムか？
  - アカウントがハックされている？
  - 患者のがんは悪性？良性？

0か 1か？

yが二つのみではない場合もある。

$$y \in \{0, 1, 2, 3\}$$



ロジスティクス回帰

0から1までの仮説がほしい。

$$h_\theta(x) = \theta^T x$$

↳ week2 の 縦ベクトル × 横ベクトル これ

これを

$$h_\theta(x) = g(\theta^T x)$$

$$g(z) = \frac{1}{1+e^{-z}}$$

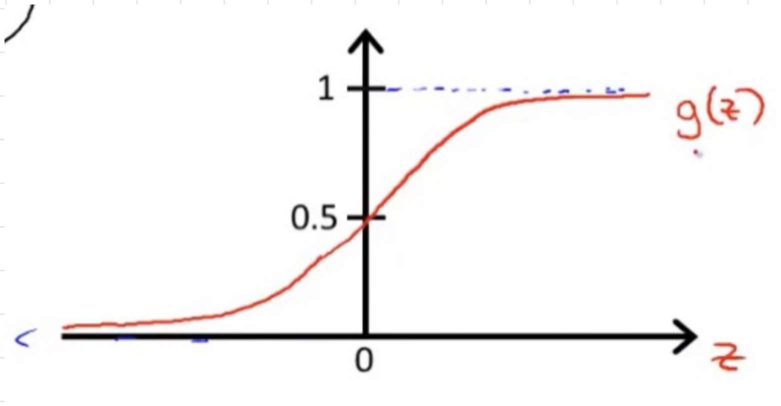
※ z は実数

g() は sigmoid 関数、または logistic 関数と呼ばれている。

これらを組み合わせて、

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

とすることができる。



パラメータθが必要になる

yは1か0にならなくてはいけない。
しかし、ロジスティクス関数からの返り値は
0から1。

$$x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{tumer-size} \end{bmatrix}$$

$$h_\theta(x) = 0.7$$

70% yは1になる。

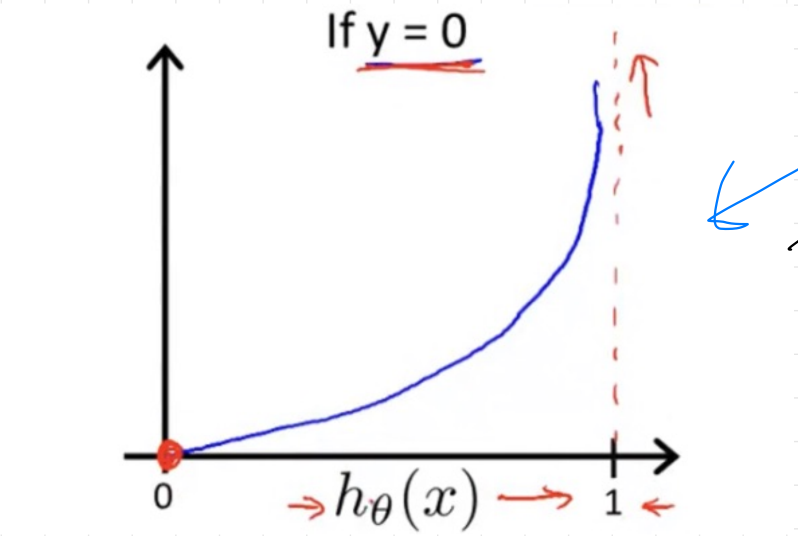
逆に

30% yは0になる。

Logistic Regressio

コスト関数　$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$

$$Cost(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

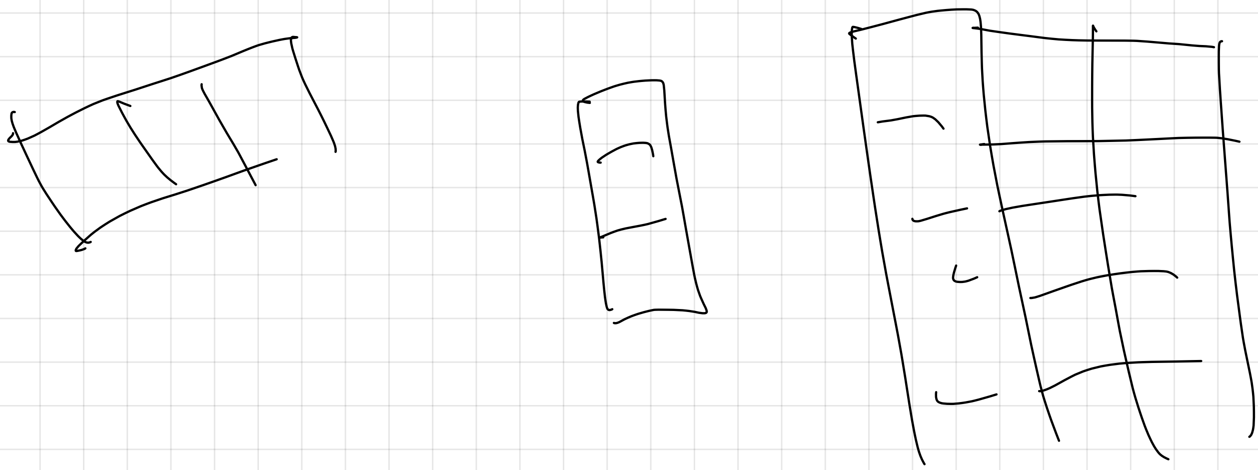


yが0のときに、
$h_\theta(x)$が1だと
予測した場合、
手痛いペナルティ

$h_\theta(x) = \infty$

分類問題だから、

yは常に0か1



Logistic regression cost function

コスト関数をコンパクトにする。

$$Cost(h_\theta(x), y) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$$

yが0か1かによって項が消える。

これを組み込んで、

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)}\log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right]$$

例によってθを最小化しなくてはいけない。

最急降下法（Gradient Descent）
を使う。

Repeat {

$$\theta_j = \theta_j - \alpha\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$

}

線形回帰に似ている…？

⇓

仮説が $h(x) = \theta^T x$ から、$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$

「におきがわっているので別物。

でも応用が効く。

（スケーリングとか…）

Advanced Optimization algorithm

- 共役勾配法
- BFGS
- L-BFGS

（

これらを学ぶには何日も、何週間も必要

大体の人がソフトウェアのライブラリを利用

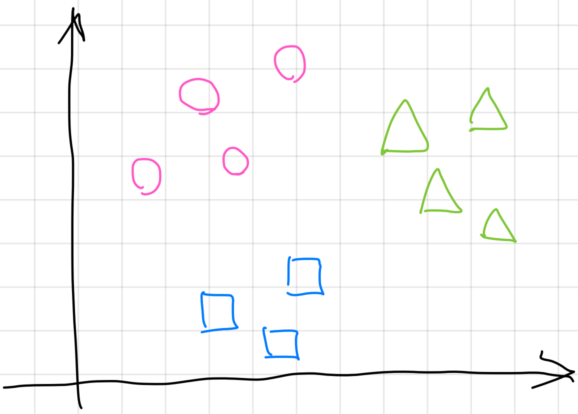
コードは

Advanced Optimization で

復習しよう。

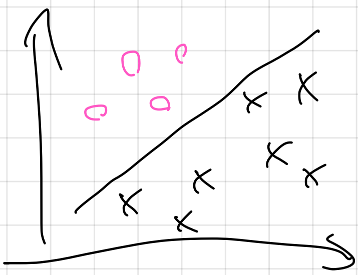
Multiclass classification

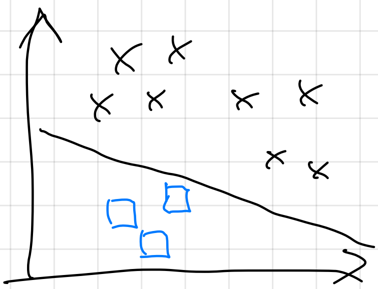
yが複数ある分類

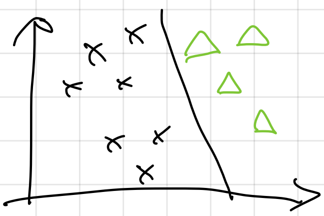


どうやって分類?

One-vs-all







$y \in \{1, 2, \ldots, k\}$

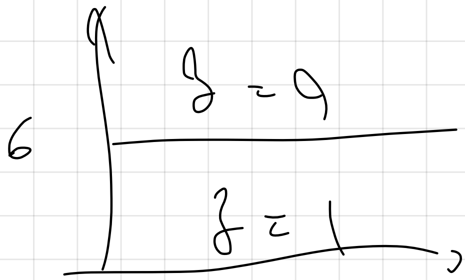
k回分類問題を行えばよい。

$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$

$\theta_0 = 6, \quad \theta_1 = 0, \quad \theta_2 = -1$

$6 - x_2$

$6 = x_2$



$-25 + 0.2x_1 + 0.2x_2 \geq 0$

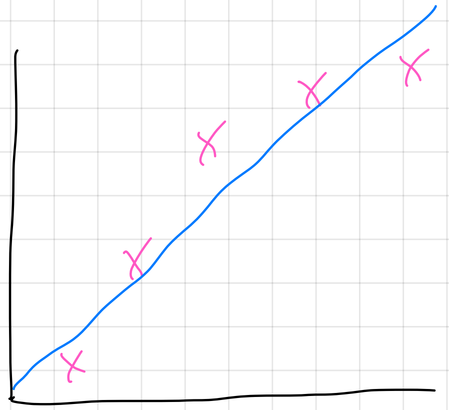
$-6 + x_1$

$0.2x_1 + 0.2x_2 \geq 25$

$x_1 = 6$

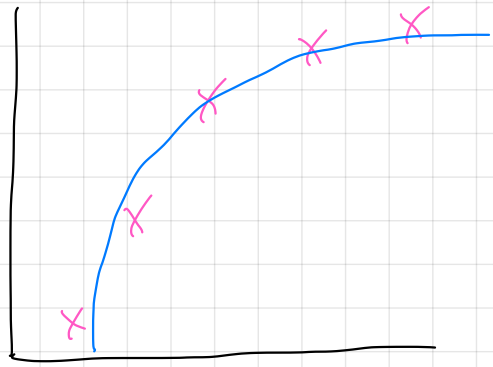
$\frac{x_1 + x_2}{0.2} = 5$

The Problem of Over fitting



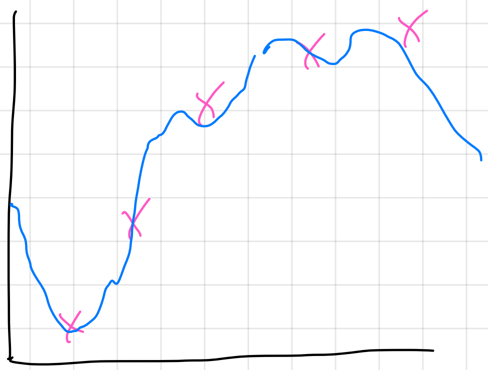
$\theta_0 + \theta_1 x$

" Under fit "



$\theta_0 + \theta_1 x + \theta_2 x^2$

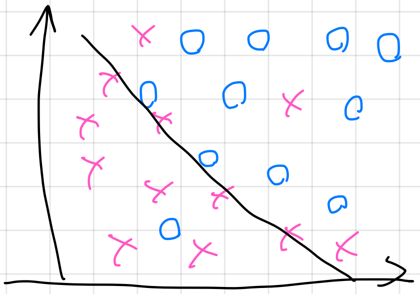
" Just right "



$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

" over fitting "

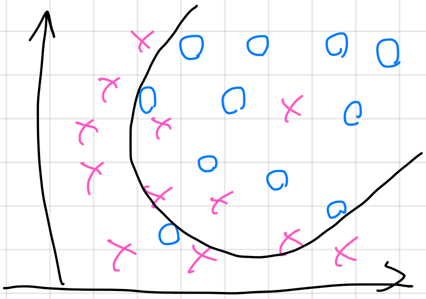
ロジスティック回帰



"Under fitting"

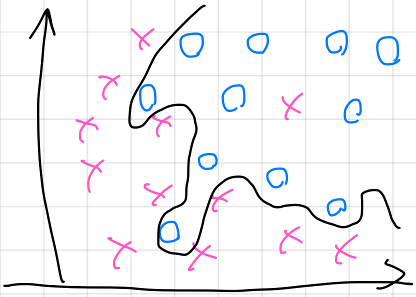
$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

(g = sigmoid function)



"Just right"

$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2$$
$$\theta_4 x_2^2 + \theta_5 x_1 x_2)$$



"Over fitting"

$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2$$
$$+ \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2 + \dots)$$

Overfitting の回避

1. 特徴数を減らす。

※ 重要な x が無くなると痛い場合がある。

2. パラメータを減らす／調整する。

※ 正則化と呼ばれる技術

正則化

θ_3 と θ_4 を 0 に近づける。

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

全てのパラメータを縮める。

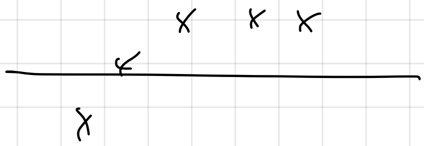
⇒ どのパラメータを短かくするべきかわからない。

$$J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 + \underbrace{\lambda}_{\text{正則化パラメータ}}\sum_{j=1}^{n}\theta_j^2\right]$$

? λが過剰に大きいとどうなる。

($\lambda = 10^{10}$ とか…)

⇒ underfitting する



※ わからない

正規化するときに、ペナルティが課されるのは、$\theta_{1, 2, 3 \ldots n}$ のみ。

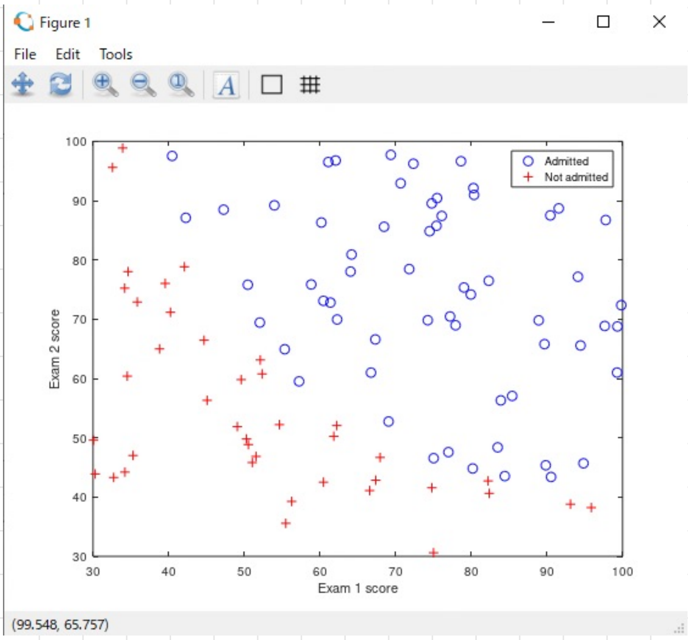
→ 0は別に用意する。

$$-25 + 0.2x + 0.2y \geq 0$$

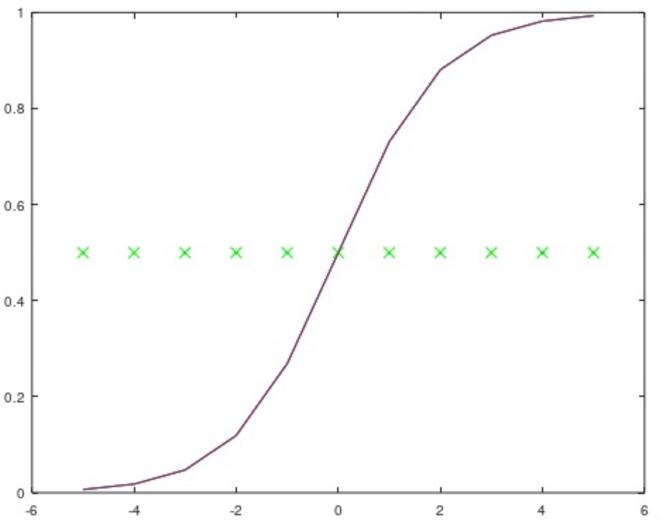
$$0.2x + 0.2y \geq 25$$

正則化を加えた
最急降下法を

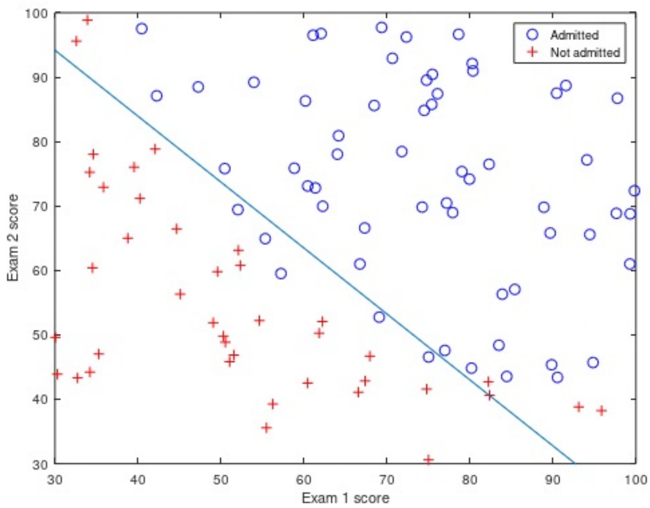
Ω を抜け



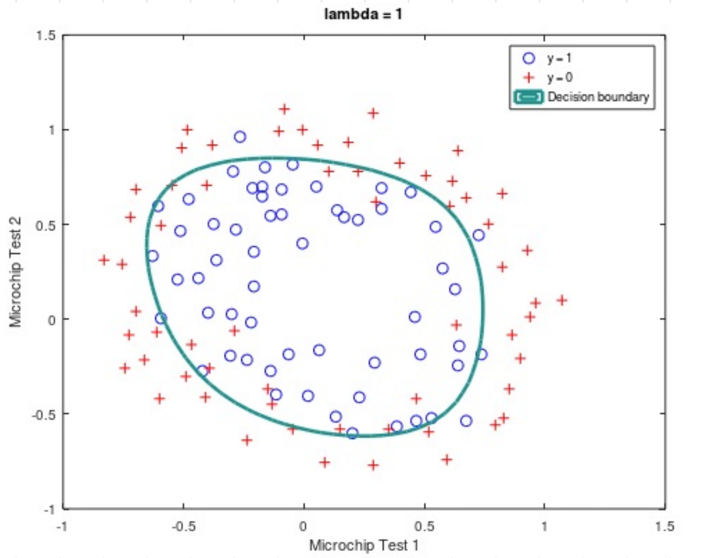
plot



sigmoid



plot



正則化つき実行

week 3

おわり