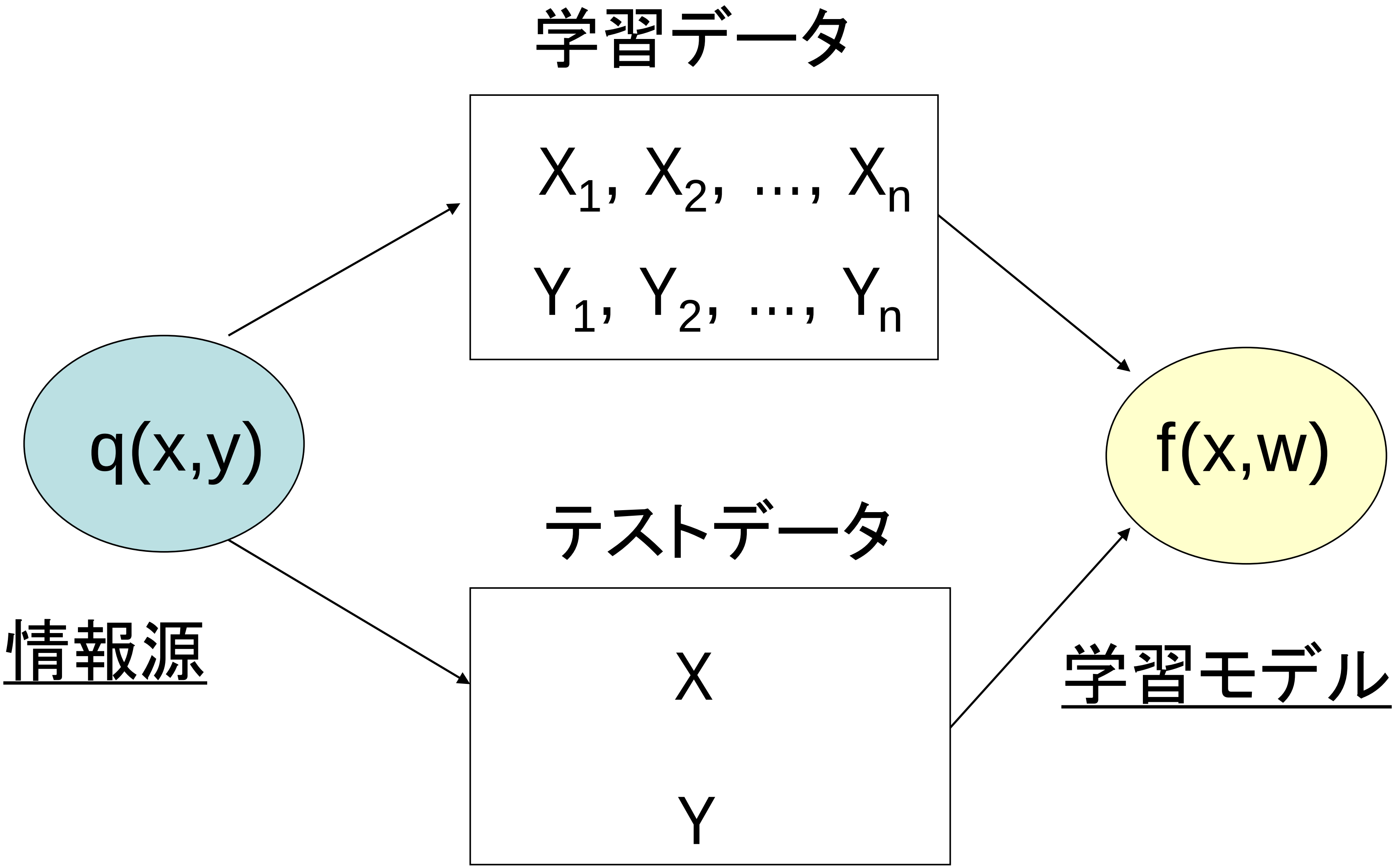
情報学習理論

渡辺澄夫
東京工業大学

教師あり学習の枠組み

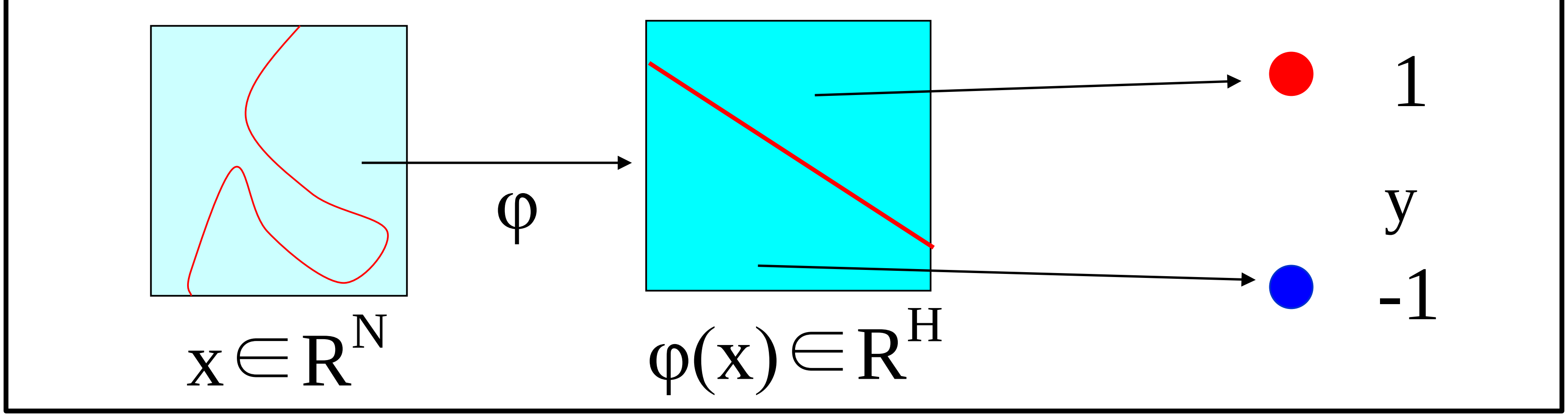


復習：SVMのしくみ

学習データ $\{ (x_i, y_i) \; ; \; i = 1, 2, \ldots, n \}$

$x_i \in R^N, \quad y_i = 1 \text{ or } -1$

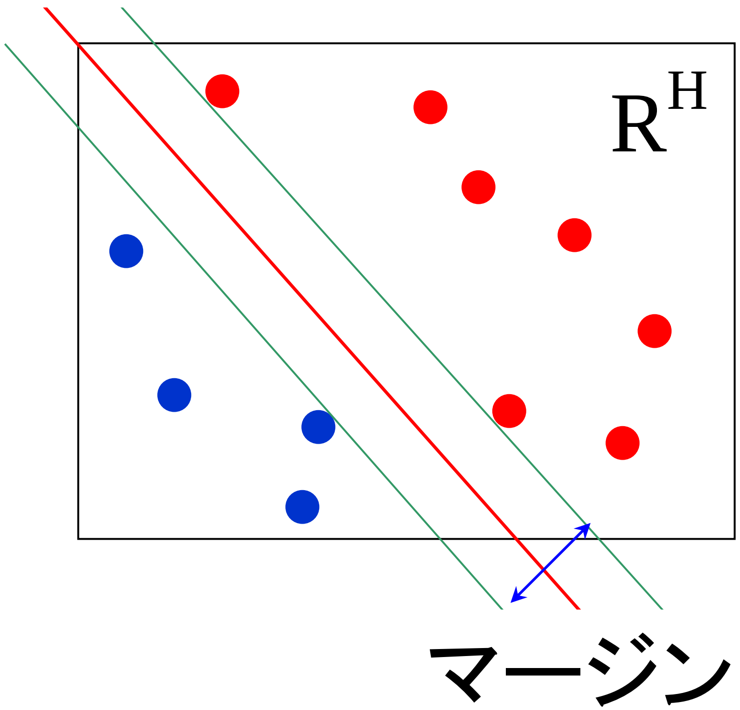
SVMの構造



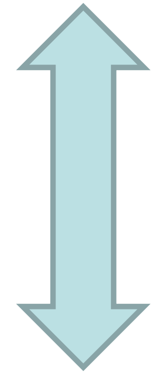
$$y = \mathrm{sgn}(w \cdot \varphi(x) + b) \qquad \mathrm{sgn}(t) = \begin{cases} 1 & (t \geqq 0) \\ -1 & (t < 0) \end{cases}$$

復習：マージン最大化と双対問題

$y = w \cdot \varphi(x) + b$



マージンを最大にする
w b を選ぶ



双対
問題

サポートベクトルを
見つける

質問：双対問題の解$\{a_i\}$はいつ大きな値になる？

正例と負例が複雑になっていて識別が難しいデータに対する
a_i は大きくなる。$0 \leqq ai \leqq C$ の制限により過剰学習を制限できる。

復習:双対問題による解

双対問題を解くことにより　最適な (w,b) はサポートベクトルに関する係数 $\{a_i\}$ を用いた和で表される

$$w = \Sigma\, a_i y_i \varphi(x_i)$$

$$b = y_i - w \cdot \varphi(x_i) \quad (a_i > 0 \text{となる任意の} i \text{について})$$

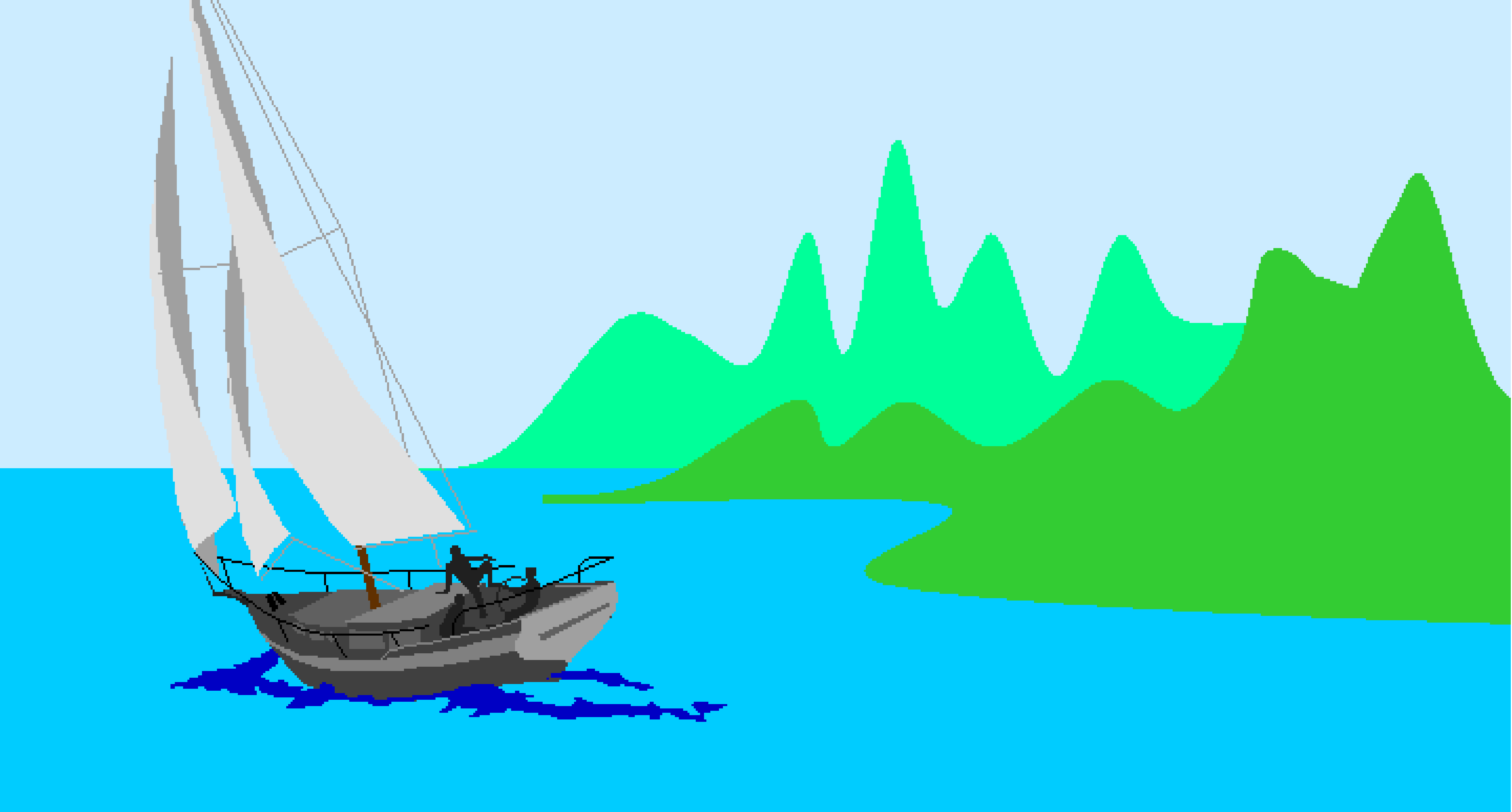
識別を行う関数は次のものになる

$$y = \mathrm{sgn}(\, \Sigma\, a_i\, y_i\, \varphi(x_i) \cdot \varphi(x) + b\ )$$

以上は、任意の関数$\varphi(x)$ に対して実行可能。

今日の課題:関数$\varphi(x)$ をどのように作る?

非線形な境界を作る



関数 φ(x)

関数 $x \in R^N \longrightarrow \varphi(x) \in R^H$

が非線形であると、非線形な境界が実現できる。

いろいろな種類の φ が提案されているが
最適な関数が何かはわかっていない。

計算量を減らすための工夫が行われている。
普通は H >> N であるため、H次元での計算を
N 次元で実行できると　とても便利である。

カーネル･トリック

多項式の例　$\varphi : R^2 \rightarrow R^6$

$$\varphi(x_1,x_2) = (1, 2^{½}x_1, 2^{½}x_2, x_1{}^2, x_2{}^2, 2^{½}x_1x_2)$$

このとき　$\varphi(x_1,x_2) \cdot \varphi(y_1,y_2)$

$$= 1 + 2x_1y_1 + 2x_2y_2 + x_1{}^2y_1{}^2 + x_2{}^2y_2{}^2 + 2x_1x_2y_1y_2$$

$$= \{ 1 + (x_1y_1 + x_2y_2) \}^2$$

$$= \{ 1 + (x_1,x_2) \cdot (y_1,y_2) \}^2$$

中間層の内積が入力層の内積で計算できることをカーネルトリックという。

この例では R^6 の内積がR^2の内積で計算できる。

SVMで必要な計算で全てトリックが使える。

$$\varphi(x) \cdot \varphi(y) = \{1 + (x \cdot y)\}^2$$

R^6 の空間の内積がR^2の内積で計算できるので
新しいデータ x についても

$$w \cdot \varphi(x) + b = \Sigma\, \alpha_i y_i \varphi(x_i) \cdot \varphi(x) + b$$

$$= \Sigma\, \alpha_i y_i \{1 + (x_i \cdot x)\}^2 + b$$

ここで Σ はサポートベクトルの和。識別は

$$y = \mathrm{sgn}(\ w \cdot \varphi(x) + b\ )$$

$$= \mathrm{sgn}(\ \Sigma\, \alpha_i y_i \{1 + (x_i \cdot x)\}^2 + b\ )$$

カーネル関数はいつ存在？

関数 $\varphi(x)$ が任意に与えられたとき、$K(x,y) = \varphi(x) \cdot \varphi(y)$ をカーネル関数という。SVMは次式で書ける。

$$y = \mathrm{sgn}\left(\Sigma\, a_i\, y_i\, K(x_i,x) + b \right)$$

逆に $K(x,y)$ が自然な条件下で $K(x,y) = \varphi(x) \cdot \varphi(y)$ を満たす関数 $\varphi(x)$ が存在する。

マーサーの定理: 次の二つの命題は同値。

(1) $K(x,y) = \varphi(x) \cdot \varphi(y)$ を満たす関数 $\varphi(x)$ が存在する。

(2) 任意の x,y について $K(x,y)=K(y,x)$ であり、
任意の関数 $f(x)$ について $\iint K(x,y)f(x)f(y)dxdy \geqq 0$。

問1

カーネル関数を持つSVMが、もともとの空間にどのような識別境界を作るかを考えてみよう。

$$\varphi(x,y) = (x^2, y^2, x)$$

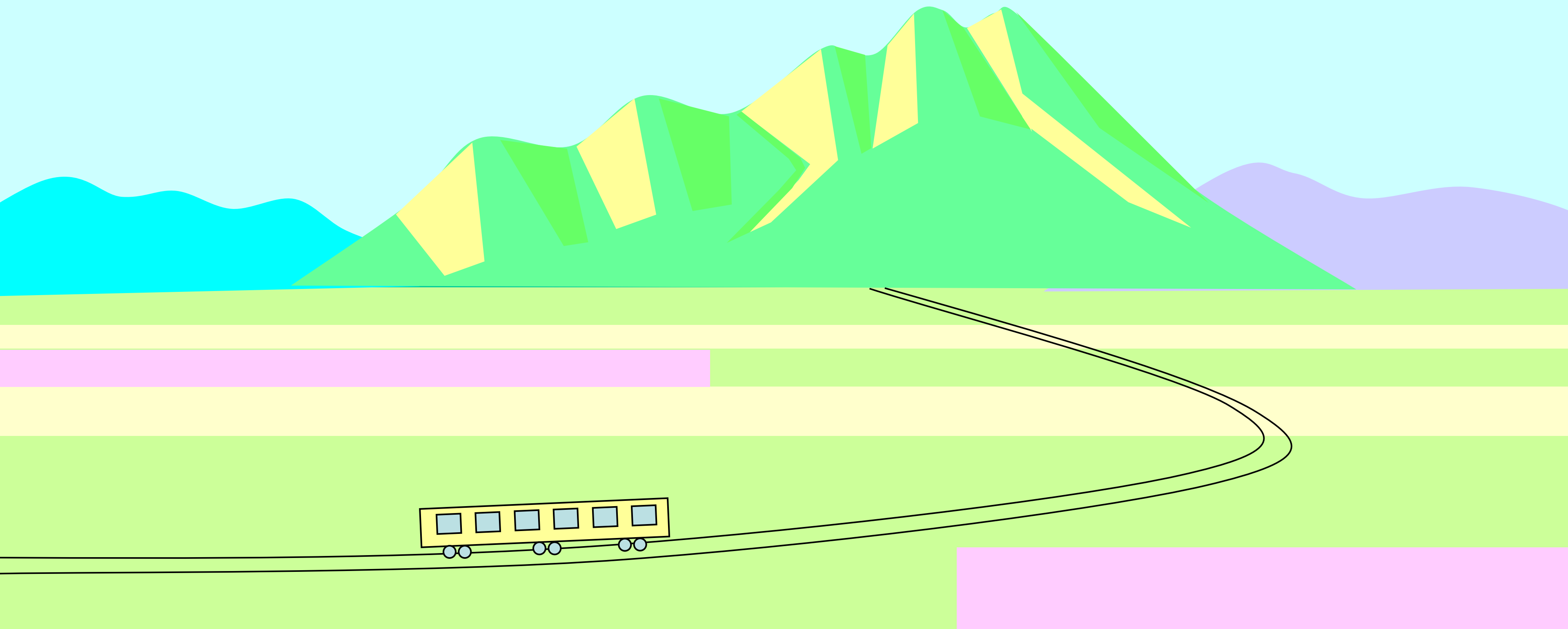
$$w = (1, a, 2)$$

とする。a が $-\infty < a < \infty$ を動くとき、集合

$$\{(x,y) \in R^2 \,;\, w \cdot \varphi(x,y) = 0\}$$

はどのように変わるか。

無限次元空間への旅



ガウスカーネル(1)

無限次元空間上のカーネルの例:　ガウス･カーネル

定数 $\beta>0$ とし、$\alpha=(2\beta)^{1/2}$ とおく。

$$R^N \xrightarrow{\varphi} R^\infty$$

$$\varphi(x) = e^{-\beta\|x\|^2}\left(1,\ \alpha x,\ \frac{\alpha x \otimes \alpha x}{2^{1/2}},\ \frac{\alpha x \otimes \alpha x \otimes \alpha x}{6^{1/2}},\ \ldots,\ \ldots\right)$$

$\otimes$ は直積を表す

ガウス・カーネル(2)

このとき

$$
\begin{cases}
\varphi(x) = e^{-\beta\|x\|^2}\left(1,\ \alpha x,\ \dfrac{\alpha x \otimes \alpha x}{2^{1/2}},\ \dfrac{\alpha x \otimes \alpha x \otimes \alpha x}{6^{1/2}},\ \dots\right) \\[2ex]
\varphi(y) = e^{-\beta\|y\|^2}\left(1,\ \alpha y,\ \dfrac{\alpha y \otimes \alpha y}{2^{1/2}},\ \dfrac{\alpha y \otimes \alpha y \otimes \alpha y}{6^{1/2}},\ \dots\right)
\end{cases}
$$

なので、内積は　($\alpha^2=2\beta$を用いて)

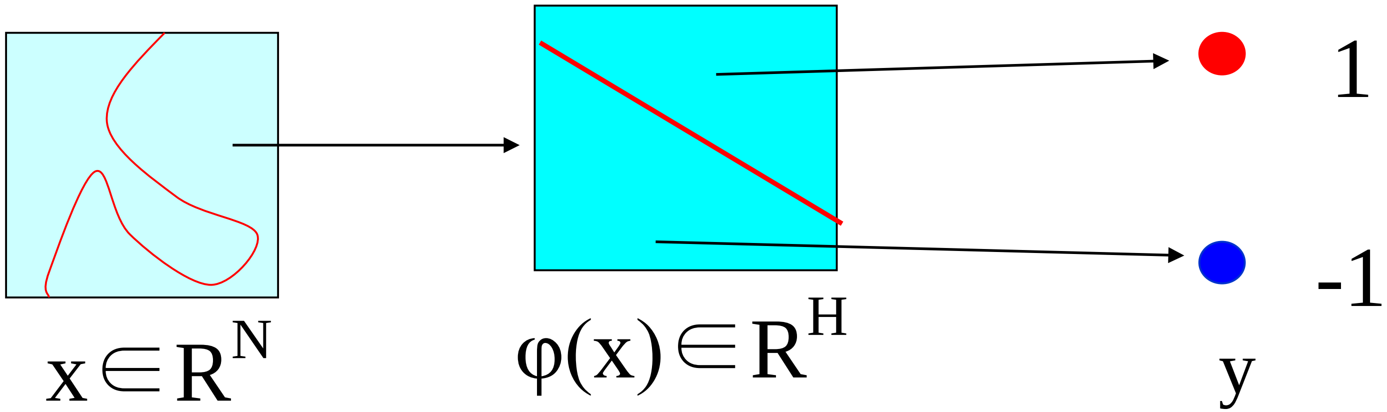
$$K(x,y) = \varphi(x)\cdot\varphi(y)$$

$$= e^{-\beta\|x\|^2}\, e^{-\beta\|y\|^2}$$

$$\times\ (1+ 2\beta\,(x\cdot y) + (2\beta\,(x\cdot y))^2/2 + (2\beta\,(x\cdot y))^3/6 + \cdots)$$

$$= \exp(-\beta\|x\|^2)\exp(-\beta\|y\|^2)\exp(2\beta x\cdot y) = \exp(-\beta\|x-y\|^2)$$

ガウス・カーネル(3)



ガウスカーネルで識別を行う関数は

$$y = \mathrm{sgn}\left(\Sigma\, a_i\, y_i\, \varphi(x_i) \cdot \varphi(x) + b \right)$$

$$= \mathrm{sgn}\left(\Sigma\, a_i\, y_i \exp\left(-\beta \|x - x_i\|^2 \right) + b \right)$$

定数 β によって識別の方法が変わる。

（注）これは、神経回路網のひとつ RBF だ。・・・

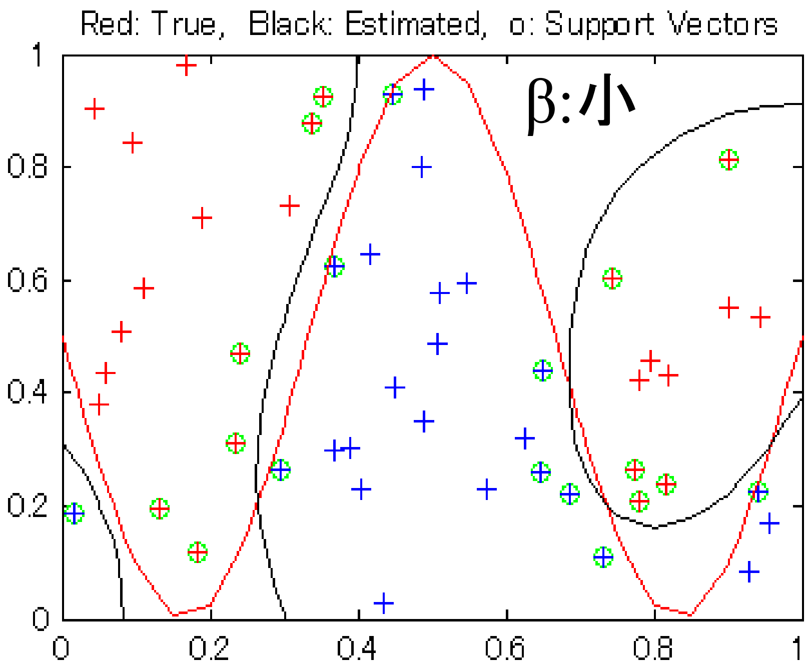
無限自由度の制御

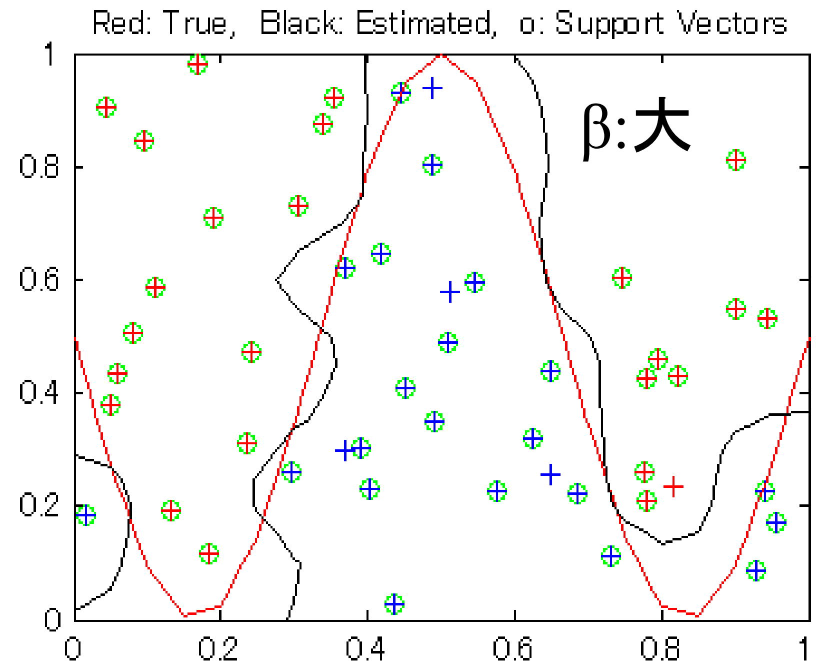
SVM　が作る関数　$y = \mathrm{sgn}(\Sigma\, a_i\, y_i \exp(-\beta\|x-x_i\|^2) + b\ )$

β:大　新しい x に対する出力は一番近い x_i だけで決まる。

β:小　新しい x に対する出力は広い { x_i }から決まる。

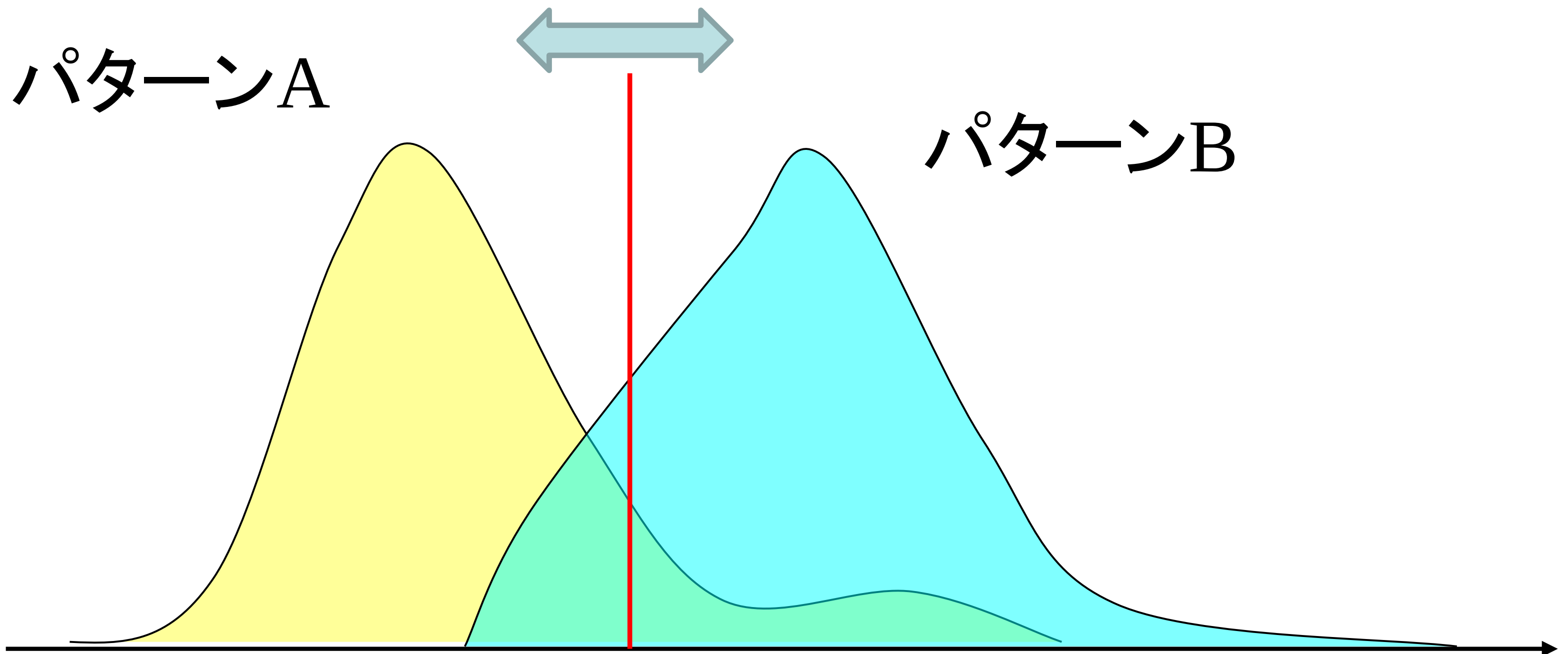
汎化誤差を小さくするには無限自由度の制御が必要である。





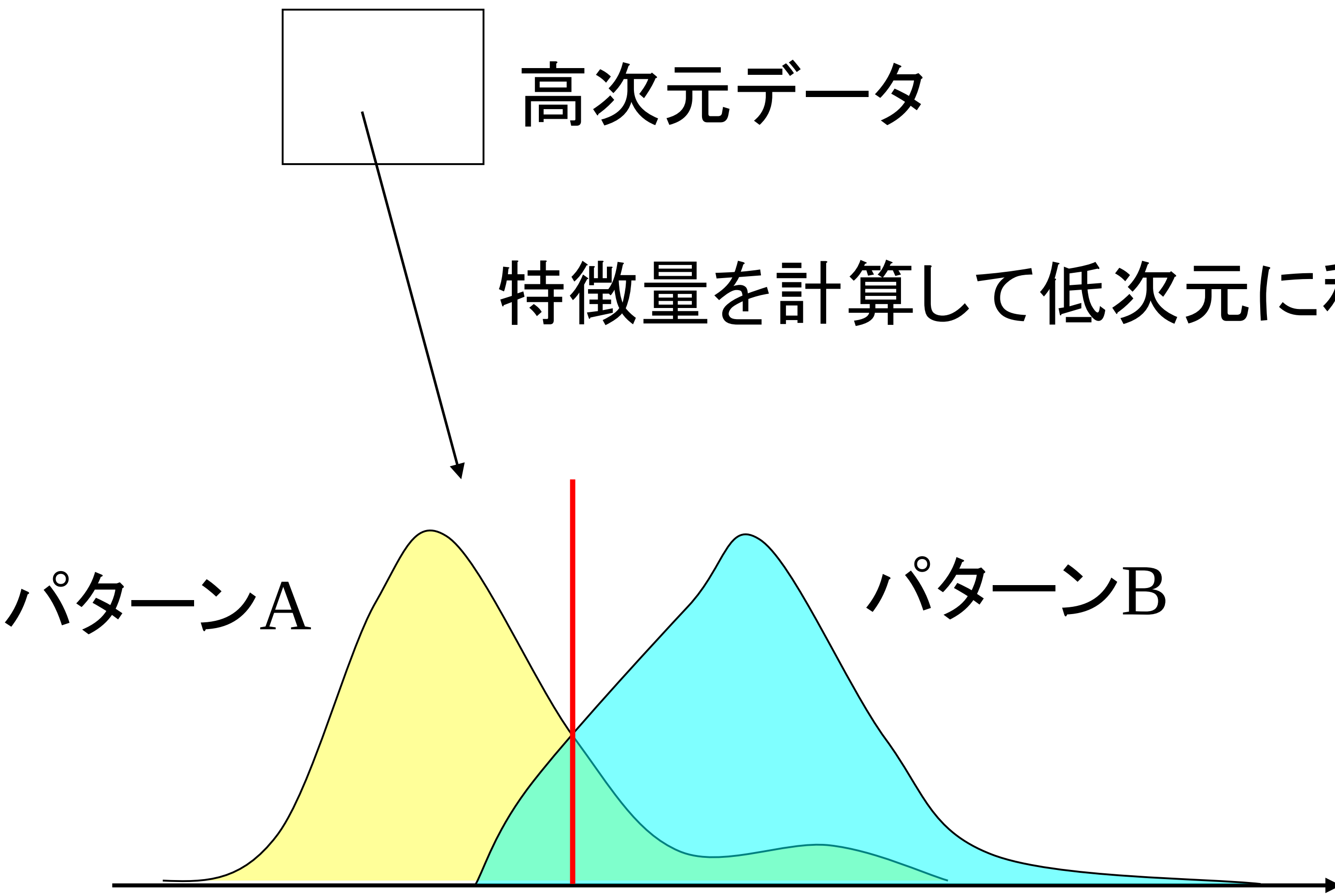
パターン識別:今昔物語(1)

もしも二つのパターンの確率分布がわかっていれば
誤識別を最小にする境界を見つけるのは難しくない。



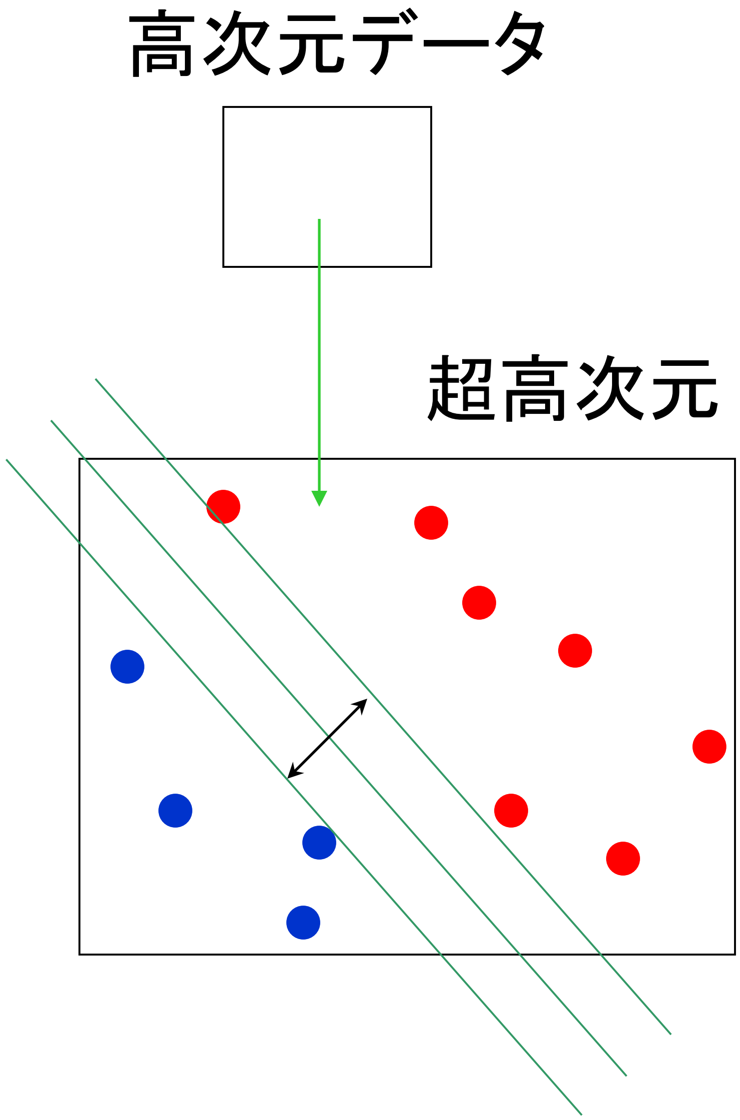
遠い昔、パターンの識別にはA,Bの確率分布を使う
方法がよいだろうと思われていた。

パターン識別：今昔物語（２）



昔、優れた特徴量を探し出す方法で競争があった。

パターン識別：現代の方法



高次元データをさらに
超高次元へ移して
線形分離できるようにする

マージン最大化により
識別境界を作る。

最適な識別ではないが
実問題で極めて強力。

人手による特徴量発見を
越えることもある。

SVMの応用

文字・画像の認識

遺伝子解析

WWWページの識別

応用多数

学習する対象についての知識を持たなくても
知識を持つ場合と同等以上の精度が実現される
こともある。SVMを使えるソフトウエアも充実。
現代のパターン識別の標準方法のひとつになった。

カーネル法の発展

SVMの成功に基づき、超高次元空間上で情報解析を行なう型の学習モデルが広く研究されるようになった。カーネルマシンという。

（カーネル回帰）実数値データ (x_i,y_i) に対して、関数

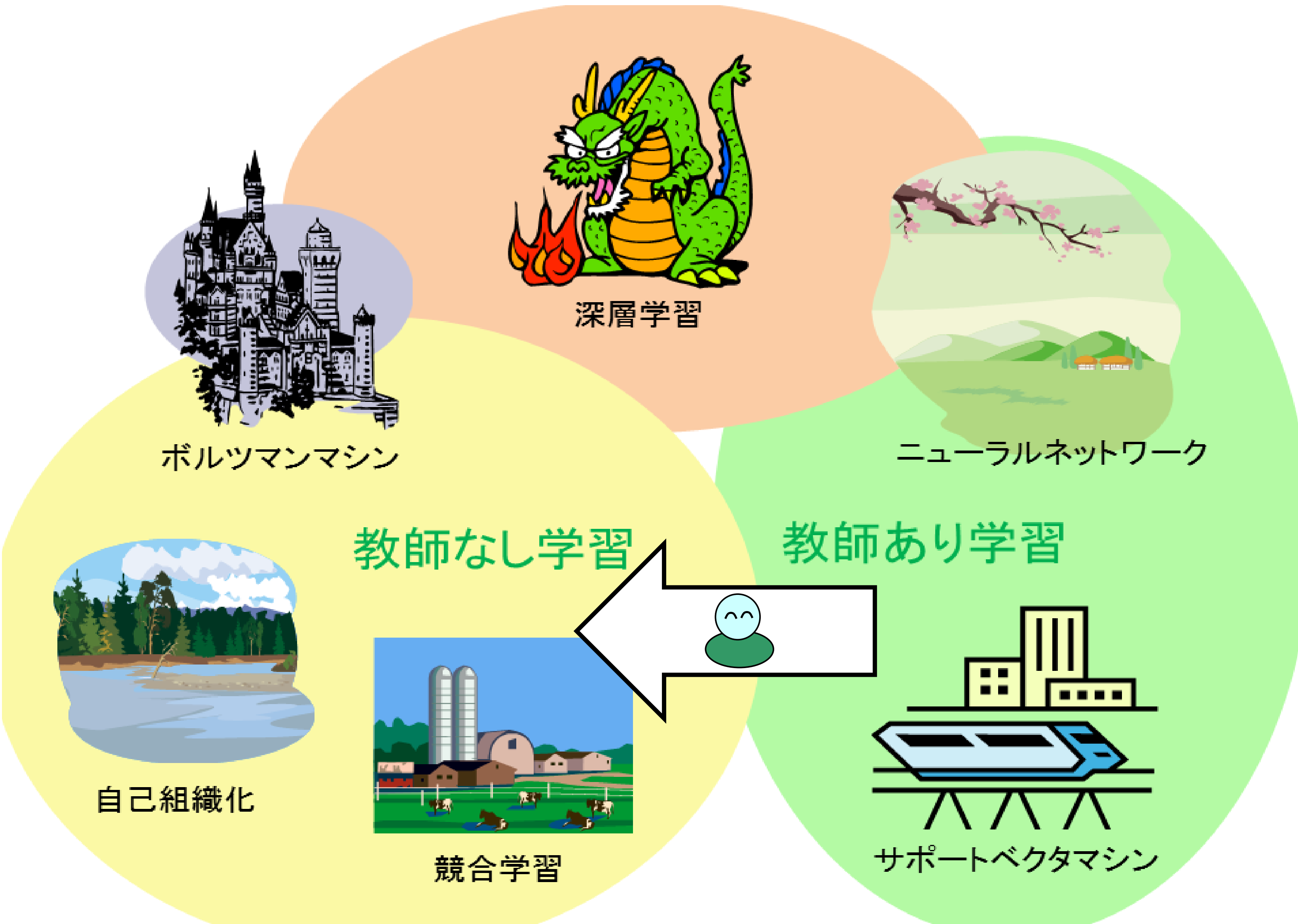
$$f(x,a,b) = \Sigma\, a_i\, y_i\, K(x,x_i) + b$$

を使って x から y を推測するために (a,b) を定める。

カーネル法による新しい統計的方法が開拓されている。

大自由度の制御法が成功のカギである。

次週予告:「人生は教師なし学習」



問2

ガウスカーネルで、カーネル関数の違いが識別境界に与える影響を調べてみよう。

βの大きさ	1	5	25	125
サポートベクトル数				
識別境界の複雑さ				
未知データ正答率				