

知能システム開発特論: 第4回

データマイニング:
分類

ダビド (david@iwate-pu.ac.jp)

IPU

22 December 2016

データマイニング は機械学習か？

MACHINE LEARNING

Academy Publish

- 分類器
- クラスタ器
- 回帰器
- ニュラルネット
- PCA
- 決定木器
- ...

知識を獲得したい！！！！



学習の手法

- 情報から従来使えそうな知識をどんな方法で見つけ出す
- 例題(事例)からアルゴリズムに基づいて概念記述(structural descriptions)を取得
- 概念記述には明示的にパターンを定義される

データマイニングの機能

● 分類と予測

- 予測を目的として、クラスや概念を区別するための記述を求める。
- 例国を気候により分類する。自動車を燃費で分類する。
- 表現: 決定木、分類規則、ニューラルネットワーク
- 予測: 未知の、あるいは欠落した数値(missing values)を予測する。

● クラスター分析

- 分類カテゴリが未知: 新しいクラスを作るためにグループ化する。
たとえば、分布パターンを見つけるために顧客の住居をクラスター化する。
- クラスターリングの原理: クラス内の類似性を最大化し、同時に、クラス間の類似性を最小にする。

回帰分析

- 独立変数と従属変数の間の関係を表す式を統計的手法によって推計する

線形回帰

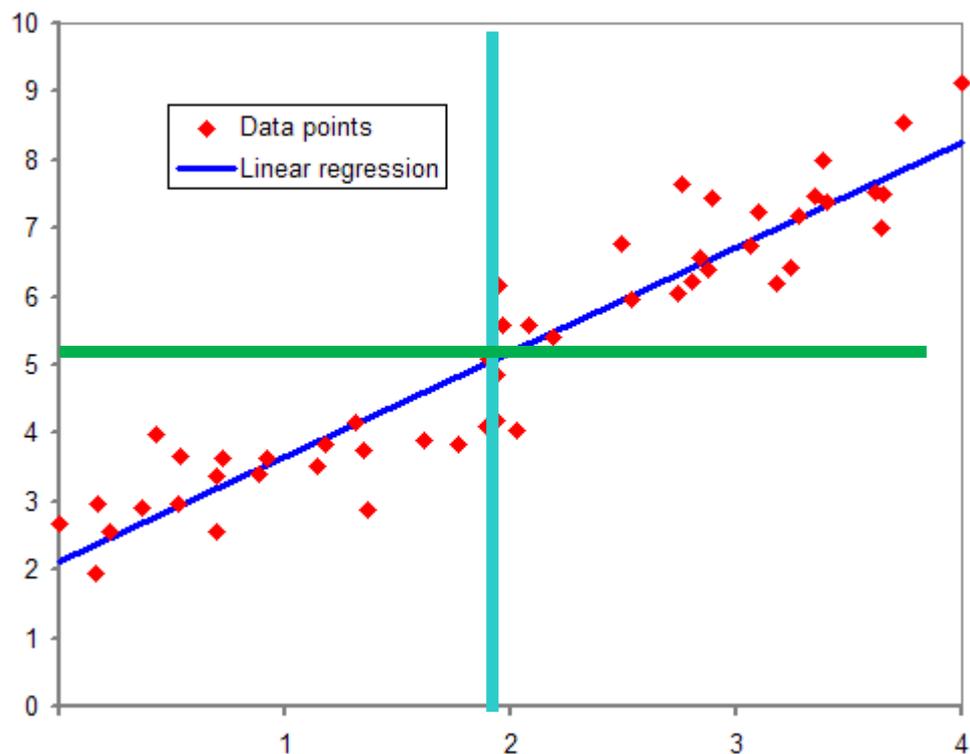
$$\hat{y} = b_0 + b_1 x$$

$$b_0 = y_{intercept}(x = 0)$$

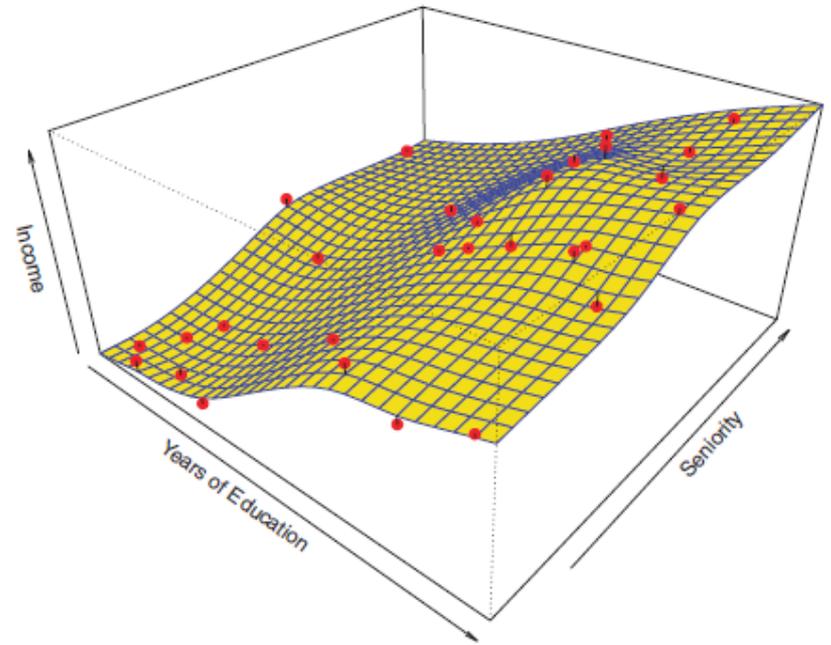
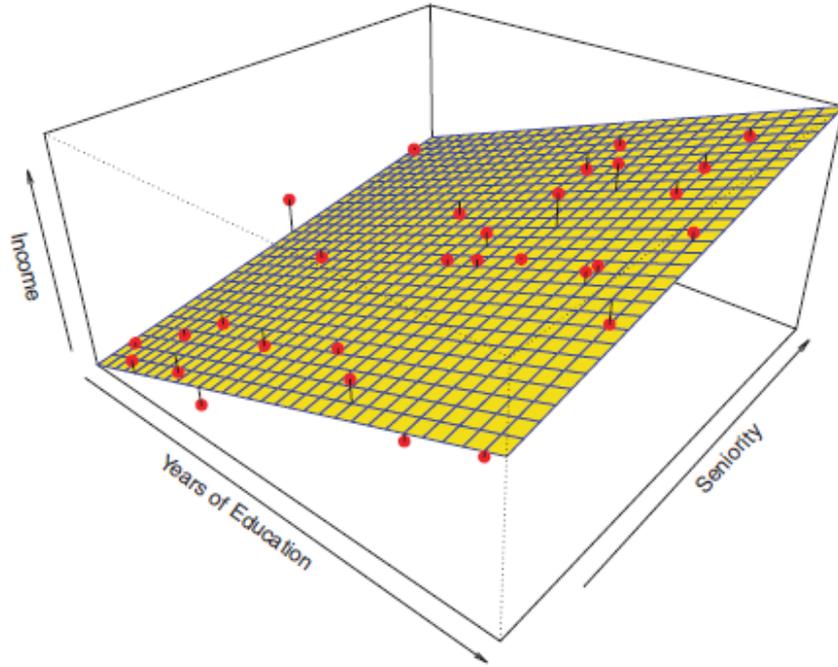
b1=線の傾き

$$b_1 = \frac{\Sigma(x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)}{\Sigma(x_i - \hat{\mu}_x)^2}$$

$$R^2 = \frac{\Sigma(\hat{y}_i - \hat{\mu}_y)^2}{\Sigma(y_i - \hat{\mu}_y)^2}, \hat{y}_i : \text{予測値}$$



パラメトリック手法 対 ノンパラメトリック手法



$$\text{income} \approx \beta_0 + \beta_1 \times \text{education} + \beta_2 \times \text{seniority}.$$

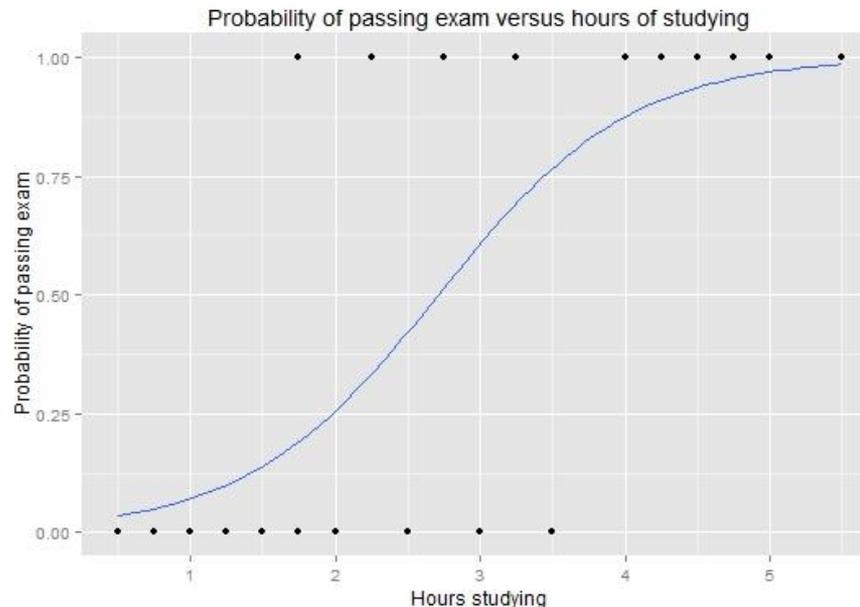
ロジスティック回帰 Logistic

regression

(https://en.wikipedia.org/wiki/Logistic_regression)

- A group of 20 students spend between 0 and 6 hours studying for an exam. How does the number of hours spent studying affect the probability that the student will pass the exam?

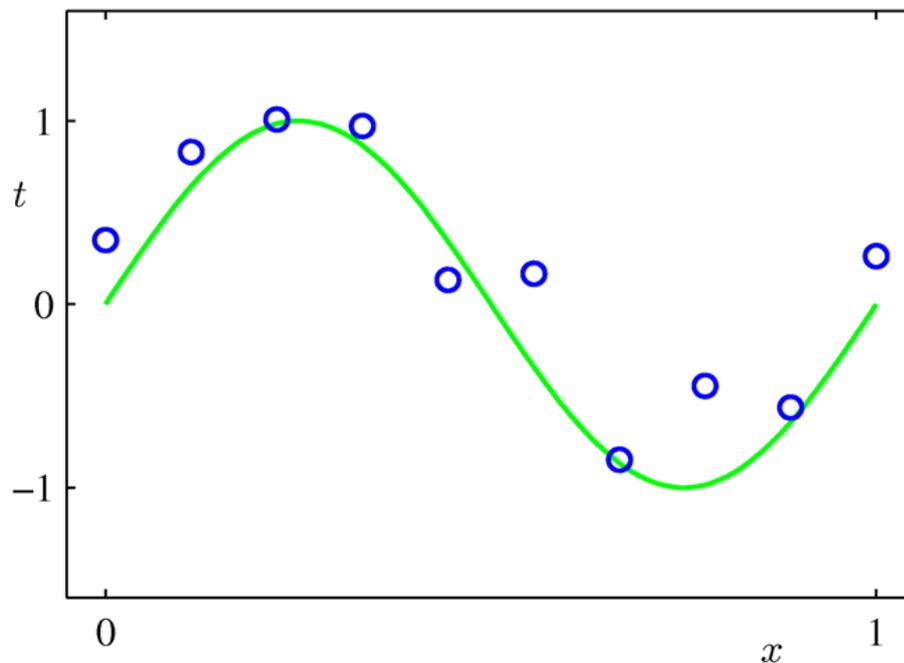
Hours	0.50	0.75	1.00	1.25	1.50	1.75	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	4.00	4.25	4.50	4.75	5.00	5.50
Pass	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	1	1	1	1	1



回帰分析

- このアルゴリズムは最小二乗法 (Least-mean-squares LMS) という名前と呼ぶ。
-
- b_1 が1に近ければ、 x と y の相関関係が強い
 - b_1 が-1に近ければ、 x と y の負の相関関係が強い
 - b_1 が0に近ければ、 x と y の相関関係がない
 - R^2 が1の近ければ、すべてのデータが回帰で説明できる。データの予測精度が高くなっている

非線形回帰



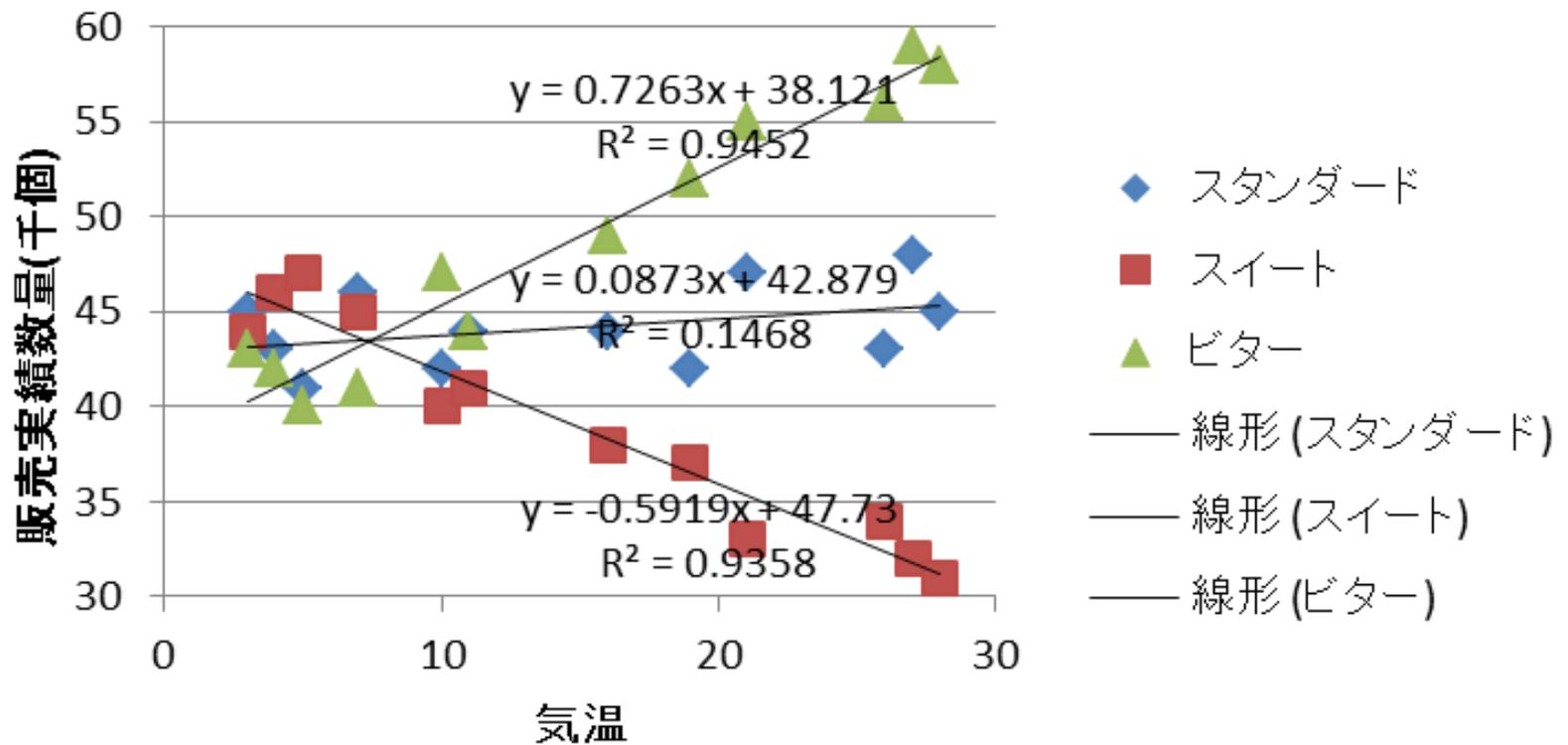
$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$$

データ例: コーヒのテストを分析する

月	昨年の実測気温	スタンダード	スイート	ビター
1	4	43	46	42
2	3	45	44	43
3	5	41	47	40
4	10	42	40	47
5	16	44	38	49
6	21	47	33	55
7	27	48	32	59
8	28	45	31	58
9	26	43	34	56
10	19	42	37	52
11	11	44	41	44
12	7	46	45	41

回帰分析の例

気温と販売数量の相関関係



気温予測に基づく販売数量

今年の予測気温	スイート	ビター
8	42.99529	43.93098
10	41.8115	45.38353
9	42.4034	44.65725
14	39.44392	48.28863
16	38.26013	49.74118
23	34.11686	54.8251
28	31.15739	58.45647
29	30.56549	59.18275
26	32.34118	57.00392
20	35.89255	52.64627
9	42.4034	44.65725
6	44.17908	42.47843

月	前年比	
	スイート	ビター
1	-7%	5%
2	-5%	6%
3	-10%	12%
4	-1%	3%
5	1%	2%
6	3%	0%
7	-3%	-1%
8	-1%	2%
9	-5%	2%
10	-3%	1%
11	3%	1%
12	-2%	4%

1年間の販売数量の予測

気温予測に基づく販売数量前年比



SVM分析

- 疎な解を持つカーネルマシン
- 最大マージン分類器
 - 重なりのあるクラス分布
 - ロジスティック回帰
- 多クラスSVM
- 回帰のためのSVM
- 計算論的学習理論

カーネル法とは？

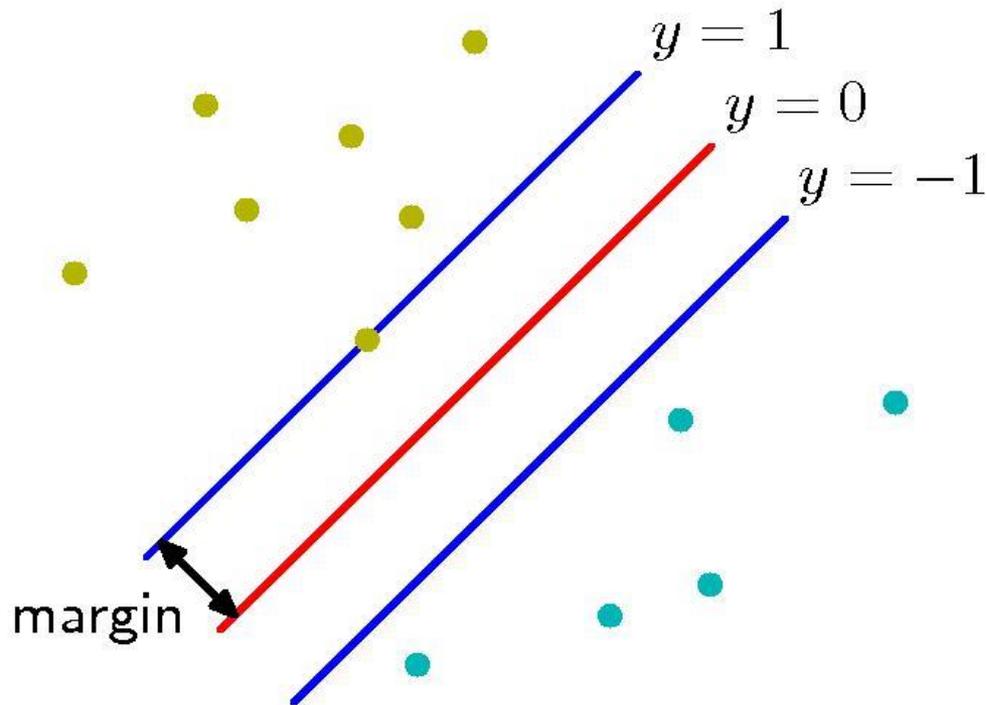
- パターン認識において使われる手法の一つ
- カーネル関数 $k(x_n, x_m)$ を利用する
 - 特徴空間中のデータの座標の明示的な計算を經由せずに特徴空間における内積をデータから直接計算できる
 - すべての訓練データ対について計算する必要があり、計算量が大きい場合が多い（=> 疎ではない解を持つカーネルマシン）

疎な解 (sparse solution) とは？

- 疎(そ)でない解を持つデータモデル
 - 学習データの大半をモデルパラメータとする
 - 例) k-近傍法
- 疎な解
 - 訓練データの一部だけに対してカーネル関数を計算することで新しい入力の予測ができる
 - 例) SVM, RVM(=> 詳しくは7.2)

最大マージン分類器

- マージンとは？



SVMでは、マージンを最大化するものを選ばれる

マージンの最大を求める

- マージンは訓練データと分類境界の最短距離であり、いま求めたいのは、そのマージンを最大化するパラメータ w と b

$$\operatorname{argmax}_{w,b} \left[\frac{1}{\|w\|} \min_n [t_n (w^T f(x_n) + b)] \right]$$

- argmax : 関数値が最大となる定義域の元の集合

マージンの最大を求める(2)

$$t_n (w^T f(x_n) + b) \geq 1$$

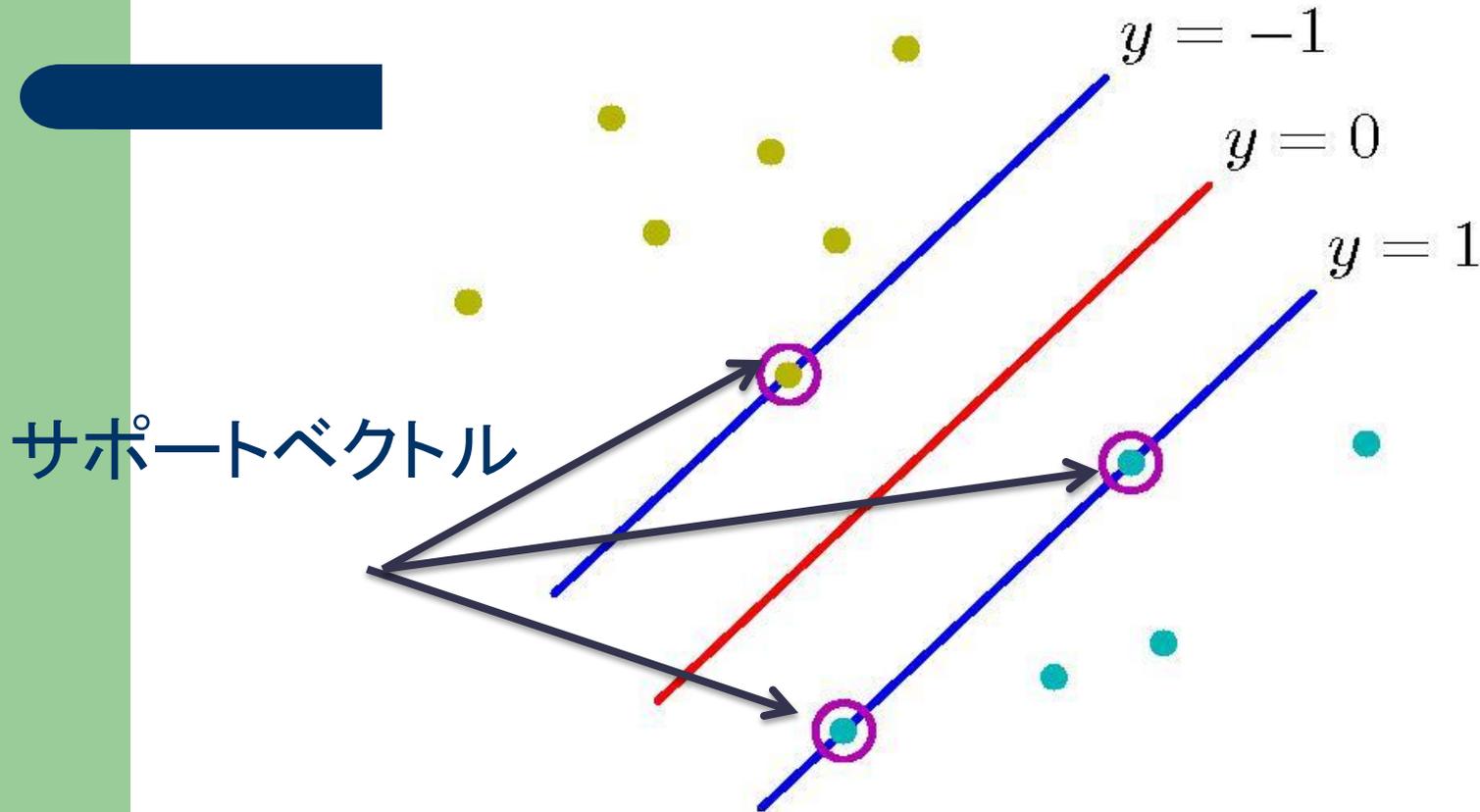
- の条件のもとで, 下記式を解くことと同等

$$\arg \min_{w, b} \frac{1}{2} \|w\|^2$$

- この式は二次計画法の一例である.
 - 線形不等式系で与えられる制約条件のもとで, 二次関数を最小化する問題

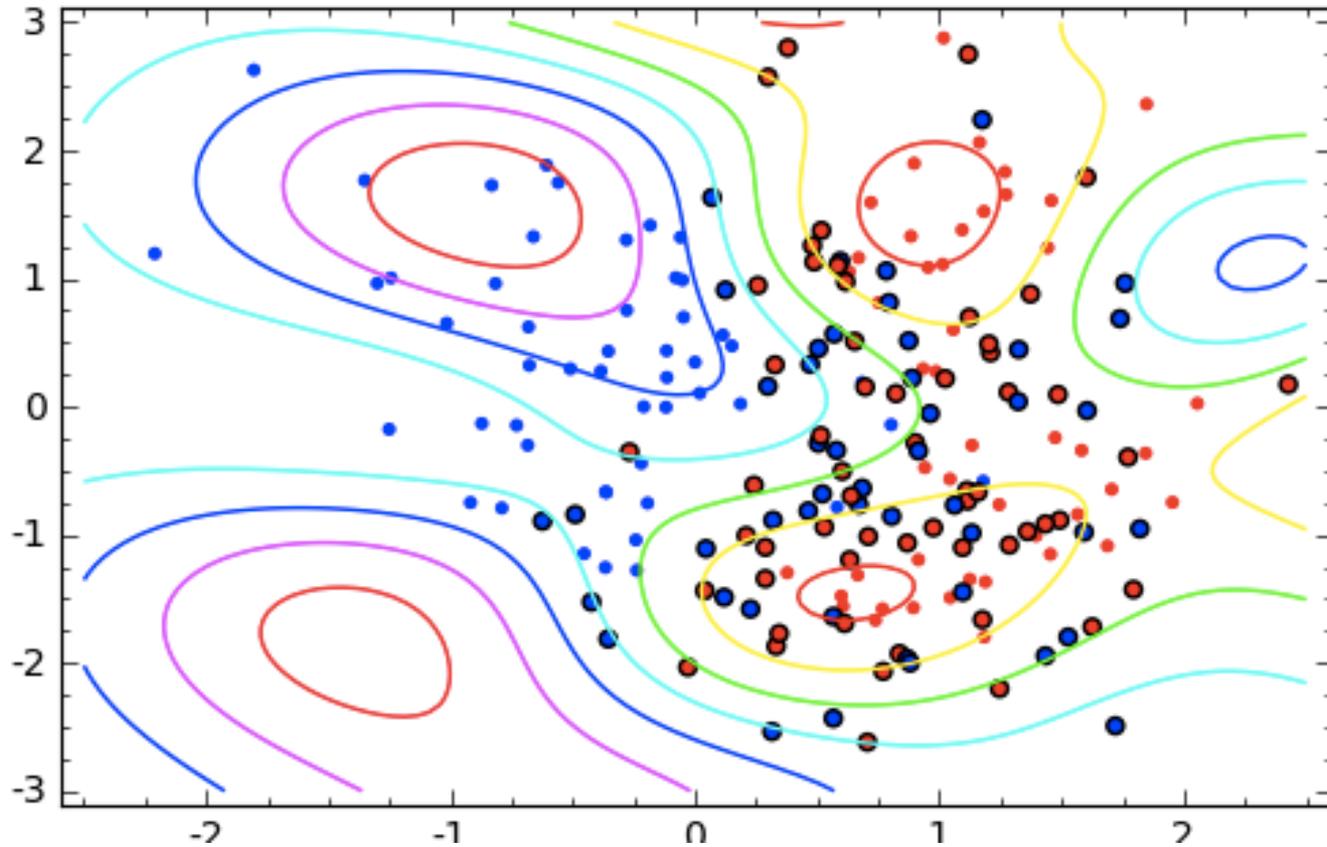
サポートベクトル

- マージンと直行する直線上の点



一度モデルを学習するとサポートベクトル以外の
訓練データは不要となる

例) ガウスカーネル関数を用いたSVM



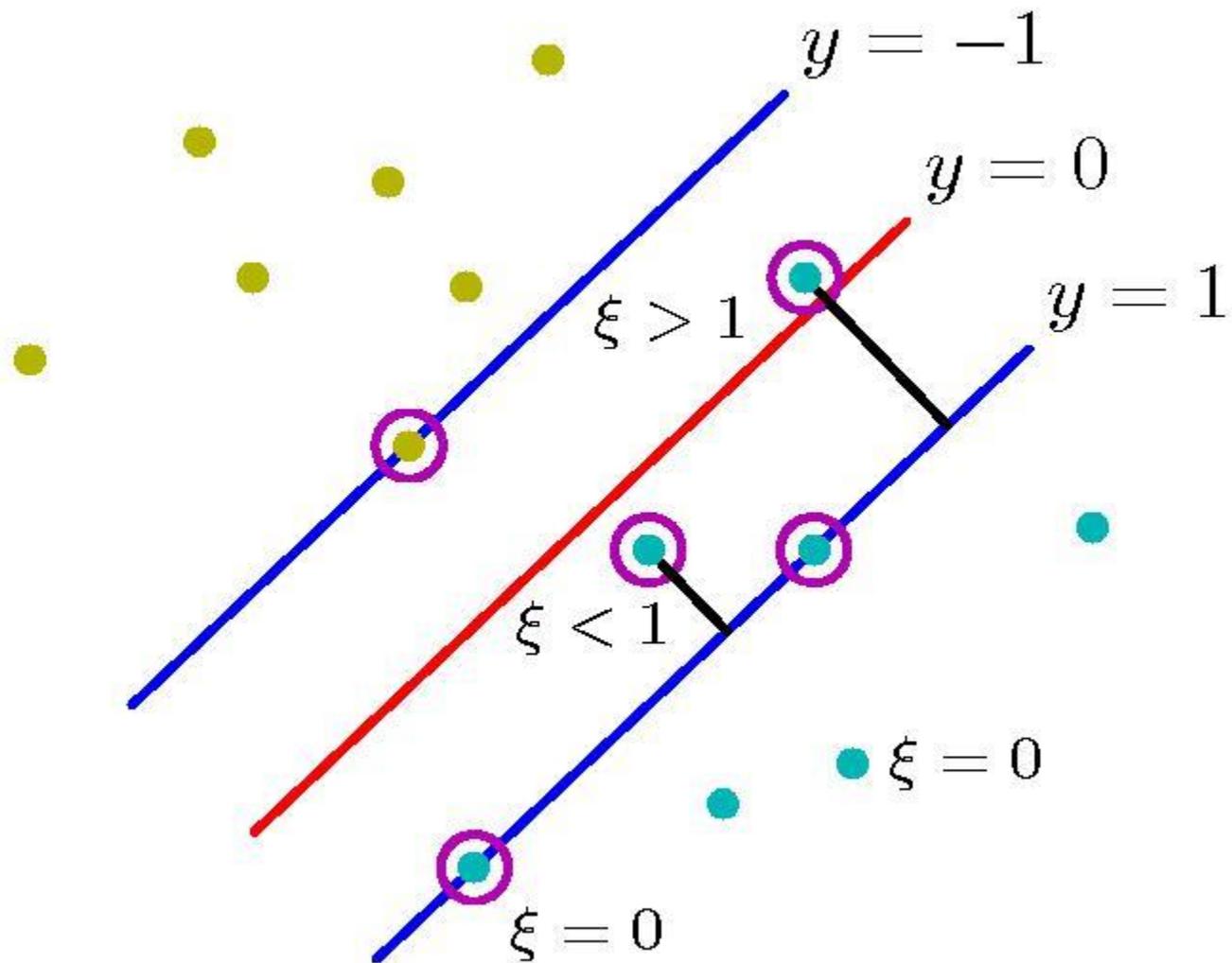
線上にある点がサポートベクトル

SVMの解は, サポートベクトル以外のデータには依存しない

重なりのあるクラス分布

- これまでの議論は、訓練データ点が完全に分離できる前提で考えてきた
- しかし、実際の問題では、完全に分離する解が必ずしも汎化能力に優れているとは限らない
 - クラスの条件付き確率分布が重なるなど
- そこで一部の訓練データに誤分類を許すようにSVMを変更する
 - スラック変数という考え方を導入する
 - 各訓練データごとに定義される変数で、データが正しく分類されかつマージン境界の上または外側に存在する場合は $\xi = 0$ それ以外の場合は、 $\xi = |tn - y(Xn)|$
- ソフトマージンSVM

重なりのあるクラス分布



多クラスSVM

- SVMは本来2クラス分類器である
 - 2クラスSVMを組み合わせることで、多クラス分類器を実現する方法が提案されている
-
- 1対他方式(よく使われている手法)
 - K個のクラスがあるときにあるクラス C_k に属するデータを正例, それ以外のデータを不例として, K個の別々のSVMを学習する方法
 - 個々のSVMによる予測矛盾が発生し, 1つの入力に複数のクラスが割り当てられる可能性
 - 1対1アプローチ
 - すべてのクラスの組み合わせについて2クラスSVMを学習し, その結果から得られた $K(K-1)/2$ 個の分類器を適用して, 多くの分類器が正例として投票したクラスを分類結果とする

多クラスSVM(2)

- DAGSVM

- 一対一方式の欠点である予測時間を削減するために, ある有効非循環グラフで表現される順番で分類器を適用する

- 誤り訂正出力符号

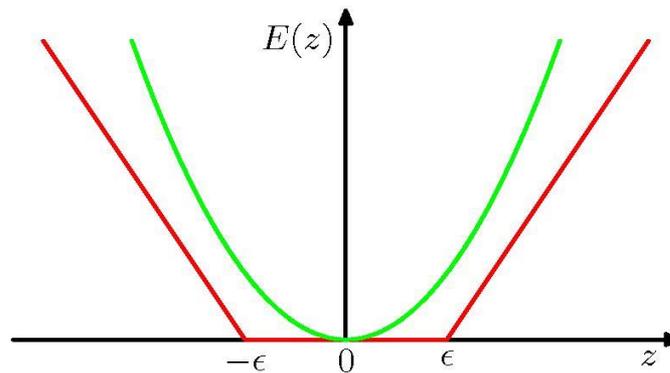
- 誤り訂正符号の原理を利用して適用する組み合わせを決定する

- 単一クラスSVM

- 確立密度分布の推定問題と関連した教師無し学習問題を解く手法
- データの密度分布をモデル化する代わりに, 高確率でデータが観測される療育を囲むなめらかな境界を求める

回帰のためのSVM

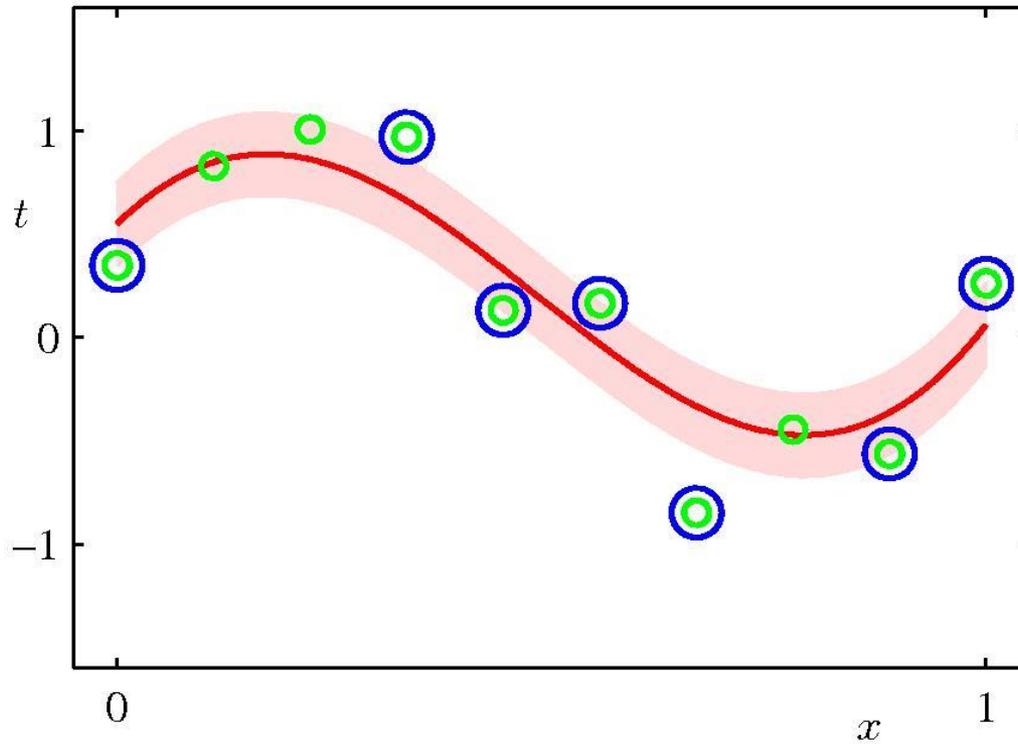
- SVMを回帰のために使うこともできる
- 単純な線形回帰問題においては，誤差関数を最小化する
 - 疎な解を得るために，二乗誤差関数を ϵ 許容誤差関数で置き換える
 - 予測値 $y(x)$ と観測値 t の差が $\epsilon (> 0)$ 未満のときは， ϵ 許容誤差関数は0となる．許容誤差外の値に関しては線形となる．



実線: ϵ 許容誤差関数
鎖線: 二乗誤差関数

回帰のためのSVM(2)

- 最適化問題をスラック変数を導入して表現することができる



回帰関数と ϵ チューブ:

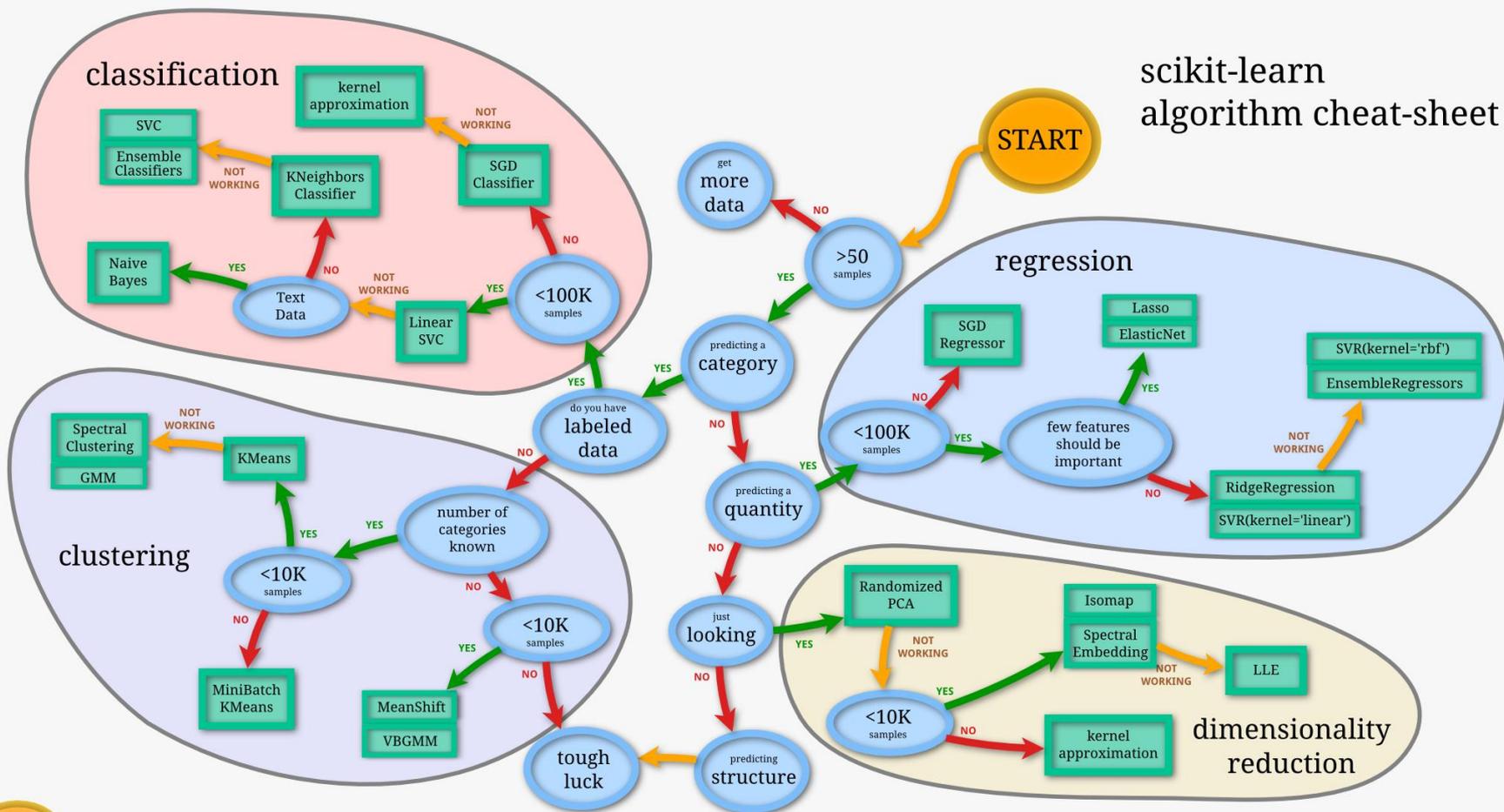
サポートベクトル(○印の点)でチューブを囲んでいる

計算論的学習理論

- SVM は、計算論的学習理論あるいは統計的学習理論と呼ばれる領域から生まれた学習アルゴリズム
-
- 計算論的学習理論は、PAC学習に端を発している
 - 良い汎化性能を達成するためにどれくらいの数の学習データが必要なのかを明らかにすることを目的とする
 - VC次元
 - 関数空間の複雑さを示す量. この概念の導入により、無限個の関数を含むような関数空間に対してもPAC学習の枠組みを適用することが可能である
 - PAC-ベイズ理論
 - PAC学習のより厳密な評価を行う取り組み

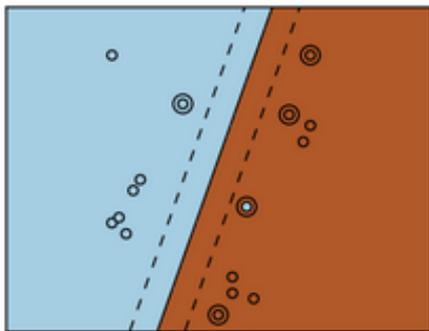
分類器/クラスタ分析器/回帰など

scikit-learn
algorithm cheat-sheet



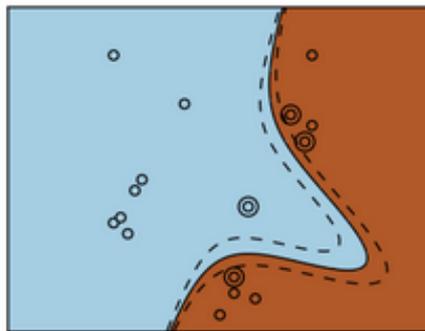
Scikit-LearnのSVMモジュール

Linear kernel



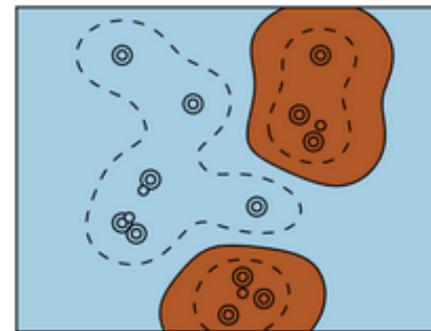
```
>>> svc =  
      svm.SVC(kernel='linear')  
      >>>
```

Polynomial kernel



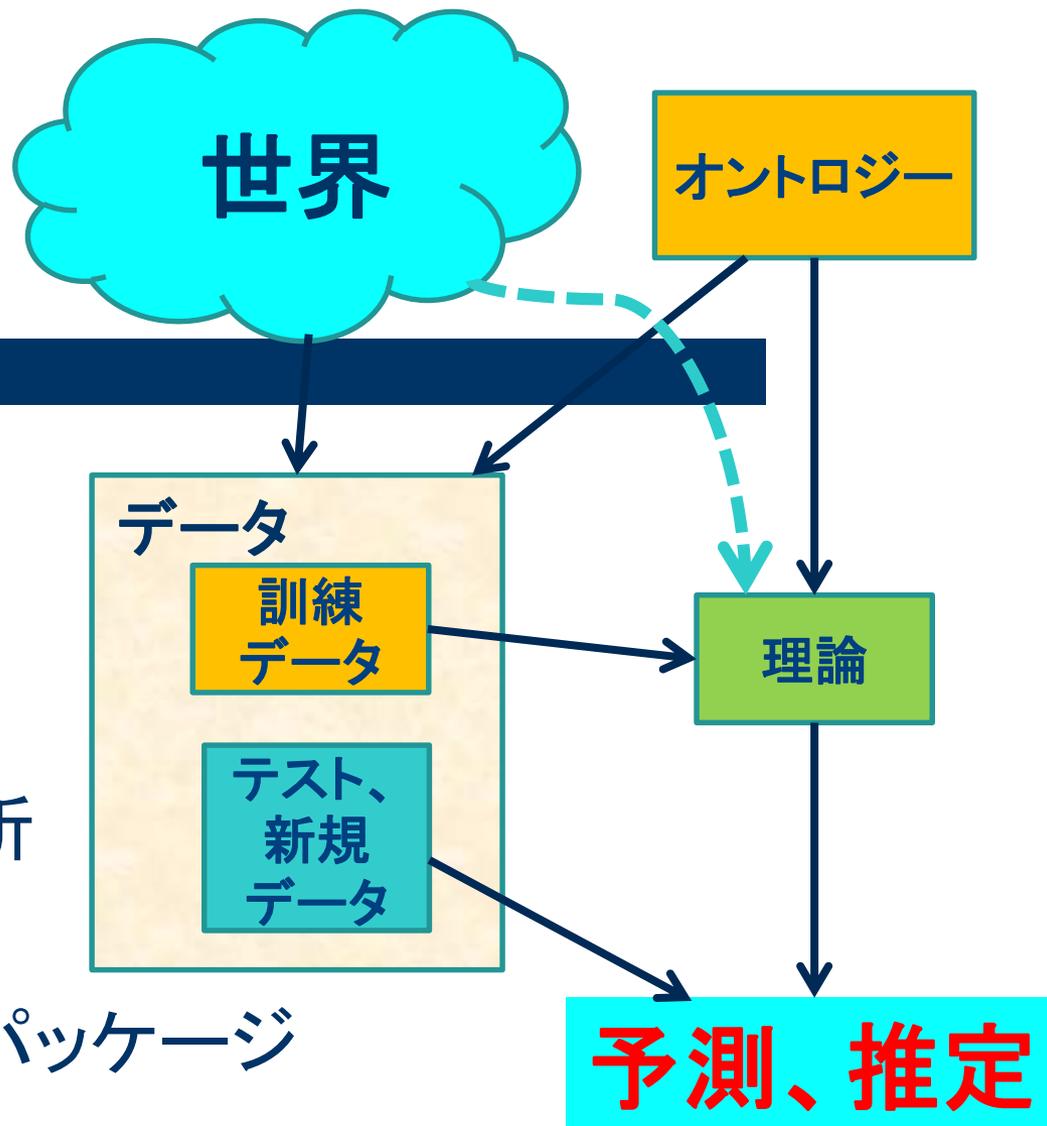
```
>>> svc =  
      svm.SVC(kernel='poly',  
              degree=3)  
      ...  
>>> # degree: polynomial  
      degree  
      >>>
```

RBF kernel (Radial Basis Function)



```
>>> svc =  
      svm.SVC(kernel='rbf')  
      >>> # gamma: inverse of  
          size of  
      >>> # radial kernel  
      >>>
```

まとめ



- 回帰分析
- SVM非線形回帰分析
- 分類例
- Python Scikit-learnパッケージ

参考文献

-
- Data Mining (Mohamed Zaki and Wagner Meira)
- "Some of the figures in this presentation are taken from "An Introduction to Statistical Learning, with applications in R" (Springer, 2013) with permission from the authors: G. James, D. Witten, T. Hastie and R. Tibshirani