

情報意味論 (課題3)

慶應義塾大学工学部
櫻井 彰人

レポート課題3

- レポートは、前回同様、web を通じて行ってください。
- レポートは電子的に作成してください。TeX, MsWord で作成して結構です。提出は pdf 形式でも結構です。
- 書くべき内容に関しては特には述べません。すべて常識的に判断してください。
- 締め切りは、2/1 23:59 とします。万が一遅れる場合には、メールにてご連絡下さい。
- 万が一の緊急の連絡のため、電子メールアドレスを書いて下さい。
- 課題1、課題2について修正・追記されたい方は、再提出して下さい。

3.1 SVM

2次元のデータを考える。
SVM分類器で、カーネル関数を使い、中心(0,0), 半径1の円の内部の点と外部の点を分ける方法を考える。

- (1) まず、非線形関数を用いて、高次元の特徴空間へ写像しよう。
この状況に適した、最も簡単な非線形関数は何か？ A hint: 3次元への写像
- (2) それに対応するカーネル関数を定めよ A hint: 上記の関数の内積
(計算が簡単になるように定める必要はない。機械的に定めて下さい)
- (3) 高次元の特徴空間に写像したあとの、分離超平面を1つあげよ。
- (4) 「中心(0,0)、半径1の円」を考えたが、これを「中心(0,0)、長軸(x軸としよう)の長さ4、短軸(y軸としよう)の長さ2の楕円であったら、円の場合と同じカーネル関数は使えるか？
- (5) 使えないなら、使えるカーネル関数は何か。

A hint: 上記(1)(2)の手順

3.2a Adaboost 簡単な証明

- Adaboost の訓練誤差の上限を証明してください。
すなわち、次の式を証明してください。

$$\text{training error}(H_{\text{final}}) \leq \exp\left(-2 \sum_i \gamma_i^2\right)$$

- 講義のスライドの式変形の部分を埋めれば結構です。当然ながら、式の個数があう必要は全くありません。
 - ヒントは、スライド中にある式です

3.2b Adaboost 簡単な実例

- Adaboost で、次のデータの学習を行って下さい。
ただし、仮説は軸並行な線によるもの(例えば、 $x > 1$ や $y < 5$)とし、2個まで用いるとします。

id	x	y	Label
1	1	10	0
2	3	11	0
3	4	4	0
4	4	2	0
5	5	10	1
6	5	6	1
7	7	7	1
8	8	8	0
9	8	7	1
10	10	14	1

- $h_1, D_1, \epsilon_1, \alpha_1, h_2, D_2, \epsilon_2, \alpha_2$ を回答して下さい。

3.3 frequent itemsets

- 次のデータに対し、Aprioriアルゴリズムを適用し k-frequent itemsets を得て下さい。
 - Minimum support は 40% として下さい。
 - 5-frequent itemsets は空になるはずです。

取引ID	トロ	酢	醤油	茶	米
1	1	1	1	0	0
2	1	1	1	1	1
3	1	0	1	1	0
4	1	0	1	1	1
5	1	1	1	1	0