

学科名	コンピュータ・ネットワーク工学	氏名	牧 隆史	指導 教員	大石 進一
研究 目	手書き文字分類に対する多クラス SVM と dPLRM の性能比較				

第1章 本論文の目的

郵便番号及び宛名自動読取区分機などに代表されるように、手書き文字を認識しクラス分類するマシンは数多く開発および実用化されてきた。

SVM(Support Vector Machine)は、Vapnik (Vapnik [1]) らによって提案された学習機械であり、汎化能力は高い。しかしクラスの分離面を決めるときに、サポートベクトルと呼ばれるデータのみしか関与せず、データの偏動はクラス分類に反映されない欠点がある。また SVM は2クラスの分類器なので3クラス以上の多クラス分類には拡張が必要となる

(Nguyen, Rajapakse [5])。dPLRM (dual Penalized Logistic Regression Machines) は、田邊 (Tanabe[6][7][8]) の多クラス分類器で、統計的な性質をクラス分類に反映し、またクラスの分類を予測確率で与える。dPLRM は多人数の会話から各々の話者を識別する実験において、GMM(Gaussian Mixture Model) の方法よりも高い性能が得られている (Matsui, Tanabe[2])。その dPLRM を用いて 0 から 9 までの手書きの数字の学習および分類を試み、マルチクラス SVM との比較を行う。

第2章 SVM(Support Vector Machine)

2クラスのデータの線形分離を考える。分離面は学習データすれすれに通るのではなく、マージンを最大にする分離面にする (Christianini, Shaw-Taylor [3])。しかしデータのノイズが多い場合、最大マージンクラス分類器では完全に線形分離は困難であるため、スラック変数 ξ を導入してマージンの制約条件を緩める。[5] 図 2.1 に概要を示す。1-ノルムソフトマージンは重みノルムとスラック変数の1-ノルムの組み合わせを最小化するになる。すなわち、

目的関数 $w^t w \rightarrow \min$

制約条件 $y_i(w^t x_i + b) \geq 1$

を解く。停留条件を考慮すると双対問題が得られる。また、カーネル関数を導入すると、非線形な分離面を構成することが可能となる。カーネル関数を導入した双対問題は次で与えられる。

目的関数

$$W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j K(x_i, x_j) \rightarrow \max$$

制約条件 $\sum_{i=1}^N y_i \alpha_i = 0, \alpha_i \geq 0$

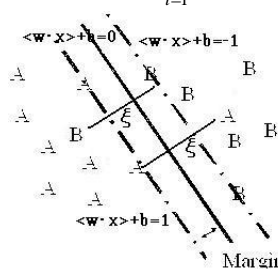


図 2.1 ソフトマージンによる分離面の構成

SVM は2クラスの分類器であったが、これを3クラス以上のマルチクラス SVM へ拡張する (Nguyen, Rajapakse[5])。原理は多クラスの判別を複数の2クラス判別に帰着させる。典型的なマルチクラス SVM の方法は one against all (OAA) である。K 個のクラス C_1, C_2, \dots, C_K の場合を考える。まず K 個の2クラス分類 SVM $C_1/\bar{C}_1, C_2/\bar{C}_2, \dots, C_K/\bar{C}_K$ を構成する。図 2.2 にアルゴリズムを示す。

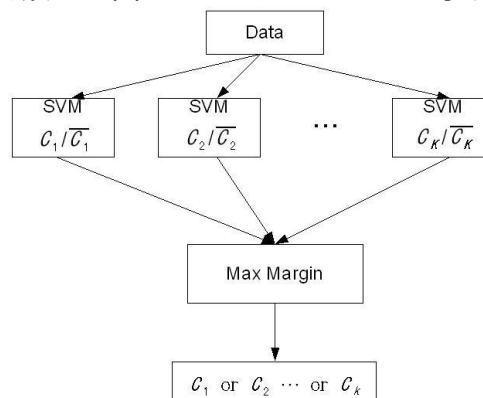


図 2.2 OAA 法によるマルチクラス SVM

あるテストデータ x のクラスはマージンが最大になるクラスに決定する。つまり、
 $\arg \max \{ \text{margin } C_j/\bar{C}_j \}$
 となるクラス C_j に決定する。

第3章 dPLRM

dPLRM(dual Penalized Logistic Regression Machines)は K 個のクラスの予測確率を与えるマシンである。予測確率の表現は確率ベクトルを用いる。

$$p(x) \equiv (p_1(x), p_2(x), \dots, p_K(x))^t \equiv \hat{p}(\overline{W} \phi(x))$$

\bar{W} はパラメータで最尤法で推定する. また過学習を回避するために罰金関数

$$P_{induct}^{\delta} \equiv \exp\left(-\frac{1}{2} \text{tr}(\Gamma \bar{W} \Sigma \bar{W}^t)\right)$$

を導入する. すると尤度関数

$$PL_{\delta}(\bar{W}) \equiv \prod_{j=1}^N \hat{p}_j(\bar{W} \phi(x_j)) \exp\left(-\frac{1}{2} \text{tr}(\Gamma \bar{W} \Sigma \bar{W}^t)\right)$$

を最大にする \bar{W} を求める. 計算を簡単にするために対数を取って負号をつけると,

$$pl_{\delta}(\bar{W}) \equiv -\sum_{j=1}^N \log \hat{p}_j(\bar{W} \phi(x_j)) + \frac{\delta}{2} \text{tr}(\Gamma \bar{W} \Sigma \bar{W}^t)$$

を最小化する \bar{W} を求める問題になる. 停留条件から双対形式に置き換えることができ,

$$p\tilde{l}_{\delta}(\bar{W}) \equiv -\sum_{j=1}^N \log \hat{p}_j(V \kappa(x_j)) + \frac{\delta}{2} \text{tr}(\Gamma V \kappa V^t)$$

を最小化する V を求める問題になる. 計算方法は反復法を用いて計算することができる.

最適解 V^* が得られたとき, あるデータに対して予測確率は,

$$p^*(x) = \hat{p}^*(V^* \kappa(x))$$

で求めることができる. マップ関数 ϕ に依存することなく予測をすることができる.

第4章 評価実験

手書き文字分類実験においてマルチクラス SVM と dPLRM による方法の性能を比較する. マルチクラス SVM は Thorsten Joachims の SVM^{multiclass} のプログラムを用いた(Joachims[4]). ソースは C 言語であり, コンパイラは gcc3.4.4 を用いた. dPLRM は MATLAB7.2 を用いて計算した. CPU は Pentium 4, 動作周波数は 2.80GHz である.

文字のデータは USPS (アメリカ郵政サービス, The United States Postal Service) が提供するデータ ([12]) を用いる. 手書き文字のデータの集合は 7291 個のトレーニングデータと 2007 個のテストデータからなる. そこから 0 から 9 までの数字を 200 個ずつ使用する. このうち 100 個ずつ計 1000 個のデータをトレーニング用, 残りの 100 個ずつ計 1000 個のデータはテスト用として用いる.

マルチクラス SVM では, 1-ノルムソフトマージンを用いて分類した. パラメータは $C=1.0$, カーネル関数は線形(linear)カーネル, 多項式(polynomial)カーネル(パラメータは $c=1$, $p=2,3$) を選択した. dPLRM では, カーネル関数はマルチクラス SVM 同様, 線形(linear)カーネル, 多項式(polynomial)カーネル(パラメータは $c=1$, $p=2,3$) を選択した. は実験的に 4^{-8} に設定した. 結果は表 4.1 のとおりである.

第5章 結論

表 4.1 を見ると, 線形カーネルを用いた場合平均

方法	カーネル	0	1	2	3	4	5	6	7	8	9	平均
SVM	linear	93	91	91	86	84	79	94	96	82	88	88.4
	2次	99	98	97	94	89	90	98	98	88	93	94.4
	3次	99	98	98	94	92	97	99	98	91	92	95.8
dPLRM	linear	91	92	91	85	84	81	91	97	80	87	87.9
	2次	99	99	96	93	90	91	97	99	88	94	94.6
	3次	99	99	99	94	92	94	99	99	91	94	96.0

表 4.1 マルチクラス SVM と dPLRM のカーネル関数と認識率

認識率はマルチクラス SVM (88.4%) の方が dPLRM (87.9%) よりも高い. しかし 2 次, 3 次カーネルでは dPLRM はマルチクラス SVM よりも高い性能が得られた. 更に 3 次カーネルを用いた場合, SVM は平均 95.8%, dPLRM は平均 96.0% の数字を正しく認識できることがわかる. しかし個々の数字を見ていくと認識の精度が低い数字も見られる. 認識に失敗した数字の多くは表 4.1 を見ると 4 と 8 は SVM, dPLRM とともに正しく認識することが難しいようである. また 5 は線形カーネルでは dPLRM の方が SVM よりも高い認識率を示しているが, 3 次のカーネルでは SVM の方が dPLRM よりも認識率は高い. ただし他の数字と比較すると線形カーネルでは 5 の認識率は低い. 一方 7 は線形, 2 次, 3 次カーネルにおいて高い認識率を示した. また SVM, dPLRM とともに認識率が高い数字と低い数字は同じ傾向を示している. 0, 1, 2, 6, 7 は認識しやすい数字といえる.

今後は学習データが少ない条件下において, マルチクラス SVM と dPLRM の性能を比較する予定である.

参考文献

- [1] Vladimir Vapnik: The Nature of Statistical Learning Theory, Springer Verlag(1995)
- [2] Tomoko Matsui and Kunio Tanabe: Speaker recognition without feature extraction process, Proc. Workshop on Statistical Modeling Approach for Speech Recognition: Beyond HMM, pp. 79-84(2004)
- [3] Nello Christianini, John Shawe-Taylor, 大北剛(訳): サポートベクターマシン入門, 共立出版社(2000)
- [4] Thorsten Joachims: Multi-Class Support Vector Machine, Cornell University(2004)
www.cs.cornell.edu/People/tj/svm_light/svm_multiclass.html
- [5] Minh N. Nguyen, Jagath C. Rajapakse: Multi-Class Support Vector Machines for Protein Secondary Structure Prediction, Genome Informatics 14, pp218-227, 2003.
- [6] Kunio Tanabe: Penalized logistic regression machines: New methods for statistical prediction 1, ISM cooperative research report 143, Estimation and Smoothing methods in Nonparametric Statistical Models, pp. 163-194(2001)
- [7] Kunio Tanabe: Penalized logistic regression machines: New methods for statistical prediction 2, Proceedings of 2001 Workshop on Information Based Induction Science IBIS2001, pp. 71-76(2001)
- [8] Kunio Tanabe: Penalized logistic regression machines and related linear numerical algebra, 京都大学数理解析研究所講究録, No. 1320, pp. 239-249(2003)
- [9] Kunio Tanabe: Private communication (2006, 2007)
- [10] B. Scholkopf, C. J. C. Burges and A. J. Smola: Advances in Kernel Methods, MIT Press(1999)
- [11] www.usps.com/
- [12] www.kernel-machines.org/