F-016

適応的な距離学習による最近傍分類器に関する一考察
A Study on Nearest Neighbor Classifier Using Adaptive Metric Learning

植田　覚 †　　松山　洋一 †　　竹内　一郎 †
Satoru Ueda　　Yoichi Matsuyama　　Ichiro Takeuchi

1　まえがき

本稿では特徴空間の距離尺度を学習するアルゴリズムである距離学習 (distancemetric learning) の問題を考察する．距離学習はおおまかに教師なし距離学習 [1, 2, 3] と教師あり距離学習 [4, 5, 6] に分類されるが, 本稿では後者を対象とする．

教師あり距離学習の主な目的は, 最近傍分類器のような事例ベース分類器の分類性能を向上させるような距離尺度を求めることである．教師あり距離学習の代表的なアプローチの一つは, 同一クラスに属するインスタンスのペアに対して *must-link* 制約を課し, 異なるクラスに属するインスタンスのペアに対して *cannot-link* 制約を課すことである [6]. この枠組では, 前者に関しては互いに距離が近く, 後者に関しては互いに距離が遠くなるように距離尺度が学習される．

一方, 最近傍分類器は近傍点の多数決によって分類結果を得るため, その判別性能は近傍点との距離のみに依存する．言い換えれば, 最近傍分類器の性能を向上させるためには, 同一クラスのインスタンスのうち近傍にあるもののみを近づけ, 異なるクラスのインスタンスのうち近傍にあるもののみを遠ざけるような距離尺度を求めればよい．本稿では, 最近傍分類性能に影響を与える近傍点を *target neighbors (TNs)* と呼び [5], これらを用いて距離学習の問題を定式化する

TNs を用いる場合に注意すべき重要な点は TNs が学習の途中で変化することである．距離学習では距離尺度自体が学習により変化するため, 学習前の距離尺度で近傍点であった TNs が学習後の距離尺度でも近傍点であるとは限らない．したがって, TNs を距離尺度に応じて適応的に変更する枠組みを持った距離学習アルゴリズムを考える必用がある．本稿では, 適応的に TNs を更新する教師あり距離学習アルゴリズムを導入し, その性能を実験的に検証する．

以下では, n 次元縦ベクトルを $\boldsymbol{v} \in \mathbb{R}^n$ のように表し, $n \times m$ 行列を $\boldsymbol{M} \in \mathbb{R}^{n \times m}$ のように表記する．また, $\mathbb{N}_n$ は 1 から n までの自然数の集合 $\{1, 2, \ldots, n\}$ を表すものとする．さらに, $\boldsymbol{I}$ は適当な次元の単位行列とする．

†名古屋工業大学, Nagoya Institute of Technology

2　問題設定と準備

本稿では, サンプルサイズ n, 特徴ベクトル次元 p の G クラス分類問題を考える． 学習データを $\mathcal{D} := \{(\boldsymbol{x}_i, y_i)\}_{i \in \mathbb{N}_n}$, $\boldsymbol{x}_i \in \mathbb{R}^p$, $y_i \in 1, \cdots, G$ とする．ここで, $\boldsymbol{x}_i$ は p 次元特徴ベクトル, y_i はクラスラベルを表す.

距離学習 (metric learning)[7, 8, 4, 5] とは, 最近傍分類器などの分類性能の向上させることを意図した機械学習アルゴリズムである．本稿では, 多くの既存の距離学習アルゴリズムと同様に, マハラノビス行列を学習する問題に限定して議論する．マハラノビス距離とは, 任意の 2 つのインスタンス $\boldsymbol{x}_i$ と $\boldsymbol{x}_j$ の距離を

$$d(\boldsymbol{x}_i, \boldsymbol{x}_j | \boldsymbol{M}) := (\boldsymbol{x}_i - \boldsymbol{x}_j)^\top \boldsymbol{M} (\boldsymbol{x}_i - \boldsymbol{x}_j) \qquad (1)$$

と表すもので, $\boldsymbol{M} \in \mathbb{R}^p \times \mathbb{R}^p$ は距離尺度を定義するマハラノビス行列である．ここで, $d(\boldsymbol{x}_i, \boldsymbol{x}_j | \boldsymbol{M})$ が距離の定義を満たすためには, $\boldsymbol{M}$ は半正定値性, および, 対称性を持たなければならない.

距離学習アルゴリズムはマハラノビス距離行列に関する最適化問題として定式化される．本稿では最近傍分類器の性能を向上させるために距離学習を行うため, 以下では, 分類誤差を損失関数として定義する．まず, 最近傍分類器の損失関数を定義するために重要な役割を担う *target neighbors (TNs)* を 2.1 節にて導入する．続いて, 2.2 節にて本稿で考察する距離学習問題の定式化を行う.

2.1　Target Neighbors (TNs)

本小節では, 最近傍分類器の損失関数を定式化するため, *target neighbors (TNs)* を導入する．説明を簡単にするため 2 クラス分類器 ($G = 2$) を考える．2 クラス最近傍分類器では, インスタンス $\boldsymbol{x}_i$ の k 個の近傍点のクラスラベルの多数決で分類を行う．したがって, 最近傍分類器の性能を左右するのは, 各クラスにおける $(k+1)/2$ 番目の近傍点であると解釈できる．例えば, $k = 3$ の 2 クラス最近傍分類では, それぞれのインスタンスにおいて, 同じクラスのうち 2 番目に近いインスタンスよりも異なるクラスのうち 2 番目に近いインスタンスが遠ければ正しく分類される (図 1 参照).

各インスタンス $\boldsymbol{x}_i, i \in \mathbb{N}_n$, において, 同じクラス $(\{\boldsymbol{x}_j\}_{j\in\{j|y_i=y_j, i\neq j\}})$ のうち, κ 番目に近いものを κ *target hit* と呼び, h_i^κ と表す. また, 異なるクラス $(\{\boldsymbol{x}_j\}_{j\in\{j|y_i\neq y_j\}})$ のうち, λ 番目に近いものを λ *target miss* と呼び, m_i^λ と表す. 以下では, target hit と target miss を合わせて, target neighbors (TNs) と呼ぶことにする. なお, TNs を定義する際には, 行列 $\boldsymbol{M}$ によって定義されるマハラノビス距離 (1) が用いられることに留意されたい.

本稿では割愛するが, 多クラス分類問題 $(G > 2)$ の最近傍分類誤差も TNs を利用して定式化可能である.

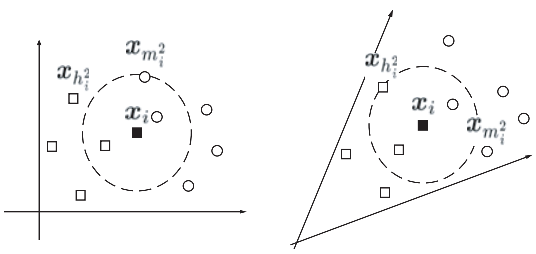

図 1: TN example

2.2 距離学習問題の定式化

最近傍分類器が高い汎化性能を持つためには, target hit までの距離よりも target miss までの距離が大きいことが望ましい. そのため, すべてのインスタンスにおいて, target hit までの距離を小さくし, target miss までの距離を大きくする距離尺度を得たい. この考え方に基づくと, 距離学習の損失関数は $n^{-1}(\sum_{i\in\mathbb{N}_n} d(\boldsymbol{x}_i, \boldsymbol{x}_{h_i^\kappa}|\boldsymbol{M}) - d(\boldsymbol{x}_i, \boldsymbol{x}_{m_i^\lambda}|\boldsymbol{M}))$ と定式化される. さらに, 正則化項を導入し, マハラノビス行列 $\boldsymbol{M}$ の制約条件を考慮すると, 本稿で考察する距離学習問題は,

$$\min_{\boldsymbol{M}} \quad \theta n^{-1} \sum_{i\in\mathbb{N}_n} (\sum_{i\in\mathbb{N}_n} d(\boldsymbol{x}_i, \boldsymbol{x}_{h_i^\kappa}|\boldsymbol{M}) - d(\boldsymbol{x}_i, \boldsymbol{x}_{m_i^\lambda}|\boldsymbol{M})) + 2^{-1}||\boldsymbol{M} - \boldsymbol{I}||_2^2 \tag{2a}$$

$$\text{s.t.} \quad \boldsymbol{M} \succeq 0, \tag{2b}$$

と定式化される. ここで, 目的関数 (2a) の第 2 項は正則化項を表し, ユークリッド距離からの乖離の程度を $\boldsymbol{M}$ と $\boldsymbol{I}$ の差のフロベニウスノルムで定量化したものである. また, $\theta \in [0, \infty)$ は正則化パラメータで, 損失関数と正則化項のトレードオフを制御するハイパーパラメータである. $\theta = 0$ のとき (2) の解はユークリッド距離の尺度となる.

3 提案法: 適応的な距離学習

既存の距離学習アルゴリズムではユークリッド距離尺度に基づいて定義された TNs が全学習プロセスにおいて固定されて利用される. TNs を固定すると, 最適化問題 (2) が半正定値計画問題となるため問題の扱いが容易となる. しかしながら, 距離学習では学習の途上で行列 $\boldsymbol{M}$ が変化するため距離尺度そのものが変更されてしまう. TNs は $\boldsymbol{M}$ により定義されるマハラノビス距離 (1) に基づいて定義されるため, 距離尺度に応じた適応的な変更が必用である. 本稿では, 適応的に TNs を変更する簡単なアプローチを提案し, その性能を実験的に検証する. 具体的には, 学習の途上で TNs の違反 (本来ならば TNs でないものが TNs となっている状況) を調べながら, その違反が大きくなった場合に, TNs をその時点のマハラノビス行列 $\boldsymbol{M}$ に基づいて更新するアプローチを検討する.

3.1 適応的距離学習の概要

以下, 提案アルゴリズムの概要を説明する. 提案法では, まず, $\boldsymbol{M} = \boldsymbol{I}$ と初期化し, ユークリッド距離尺度に基づいて初期の TNs を割り当てる. なお, ユークリッド距離尺度 $(\boldsymbol{M} = \boldsymbol{I})$ は $\theta = 0$ における (2) の解となっていることに注意されたい. 学習が進むにつれて距離尺度 $\boldsymbol{M}$ が変化すると, TNs の違反が生じてくる. 違反の程度がある基準を越えたとき[1], その時点の $\boldsymbol{M}$ において TNs の再割当を行い, さらに学習を続ける.

TNs の再割当を行う $\boldsymbol{M}$ を, 順に, $\boldsymbol{M}_0, \boldsymbol{M}_1, \boldsymbol{M}_2, \ldots$ とすると, t 回の再割当が行われた後では, 最適化問題 (2) の単位行列 $\boldsymbol{I}$ を $\boldsymbol{M}_t$ に変更した問題:

$$\min_{\boldsymbol{M}} \quad \theta n^{-1} \sum_{i\in\mathbb{N}_n} (\sum_{i\in\mathbb{N}_n} d(\boldsymbol{x}_i, \boldsymbol{x}_{h_i^\kappa}|\boldsymbol{M}) - d(\boldsymbol{x}_i, \boldsymbol{x}_{m_i^\lambda}|\boldsymbol{M})) + 2^{-1}||\boldsymbol{M} - \boldsymbol{M}_t||_2^2 \tag{3a}$$

$$\text{s.t.} \quad \boldsymbol{M} \succeq 0, \tag{3b}$$

を解く. 最適化問題 (3) の $\theta = 0$ における最適解は $\boldsymbol{M}_t$ となっている.

$t = 1, 2, \ldots$ において最適化問題 (3) を解く際には, TNs の違反が基準を越えないような最大の $\theta > 0$ を見つける必用がある. 提案法では, 簡単な二分探索を用いてこのアプローチを実装している. 具合的には, ある θ における (3) の解を求め, TNs の違反が基準を越えていなければ θ の更新幅を 2 倍とし, 基準を越えていれば θ

[1] 次節の実験では, 全 n 個のインスタンスのうち, 誤った TNs の個数が 10%以上の場合を基準としている.

の更新幅を半分にすることを繰り返すことによって, 基準を越えない最大の $\theta > 0$ の探索を行う.

なお, 次節の実験では, (3) のソルバーには, `cplex` [9] を用いた. また, 過学習を防ぐため, 学習プロセスで計算する $\boldsymbol{M}$ の系列に対する評価サンプルの最近傍分類誤差を計算しておき, それが最小となる $\boldsymbol{M}$ を選択した.

3.2 適応的距離学習アルゴリズム

適応的距離学習のアルゴリズムを以下に示す.

Algorithm 1 適応的距離学習

入力: 学習サンプル $\mathcal{D}$, 近傍パラメータ κ, λ, 初期ステップ幅 $\Delta\theta_0$;
初期化: $t \leftarrow 0$, $\boldsymbol{M}_0 \leftarrow \boldsymbol{I}$ とし, $\boldsymbol{M}_0$ に基づき, TNs $\{(h_i^\kappa, m_i^\lambda)\}_{i\in\mathbb{N}_n}$ を割り当てる.
while 終了基準を満たすまで **do**
  $u \leftarrow 0$, $\Delta\theta \leftarrow \Delta\theta_0$, $\theta_0 \leftarrow 0$;
  while $\Delta\theta$ が十分に小さくなるか θ が十分に大きくなるまで **do**
    $\theta \leftarrow \theta_u + \Delta\theta$ とし, (3) を解き, 解を $\hat{\boldsymbol{M}}$ とする
    if 現在の TNs が $\hat{\boldsymbol{M}}$ において違反の基準を下回るならば **then**
      $\theta_{u+1} \leftarrow \theta$;
      $\Delta\theta \leftarrow 2\Delta\theta$;
    else
      $\theta_{u+1} \leftarrow \theta_u$;
      $\Delta\theta \leftarrow 0.5\Delta\theta$;
    end if
    $u \leftarrow u+1$
  end while
  $\boldsymbol{M}_{t+1} \leftarrow \hat{\boldsymbol{M}}$ とする.
  $t \leftarrow t+1$
end while
出力: マハラノビス距離行列 $\boldsymbol{M}$;

4 計算機実験

本節では提案アルゴリズムのベンチマークデータに対する数値実験結果を報告する. 提案アルゴリズムと TNs を変更しない場合との最近傍分類の分類誤差を比較する. データは UCI Machine Learning Repository から入手したものを用いた. 表 1 に実験で用いたデータの概要を示す. 各データをランダムに学習データ, 評価データ, テストデータに分割する. データ数が 300 以下の場合には, データを学習データと評価データとテストデータに 3 分割し, それ以上の場合には, 学習データ数と評価データ数を 100 とし, 残りをテストデータとした. ランダムな影響を軽減するため, データの分割を 10 通り行い, その時の分類誤差の平均を表 2 に示す. TNs を変更しない場合では θ の更新幅を一定とし, 提案アルゴリズムと同様に評価データの最近傍分類誤差が最小となる $\boldsymbol{M}$ を選択した. 終了条件は提案アルゴリズム, TNs を変更しない場合共に, 損失関数の変化が十分に小さくなったときとした. 提案アルゴリズムの k は評価・テスト時の最近傍分類の近傍数と等しくなるように設定した. 表 2 は $k = 5$ の場合に選択された $\boldsymbol{M}$ をテストデータに用いた時の最近傍分類の分類誤差の平均と標準偏差を示している. 提案アルゴリズムは状況によって提案アルゴリズムが有効に働くことがあるということを示した.

表 1: データの概要

データ名	データ数	次元数	クラス数
BreastCancerDiagnostic	569	30	2
MAGICGammaTelescope	19020	10	2
Parkinsons	195	22	2
SPECTFHeart	267	44	2

5 おわりに

本稿では最近傍分類器のための適応的距離学習アルゴリズムを提案した. 最近傍分類器の損失関数を定義するために TNs を導入し, 距離学習問題の定式化を行った. 適応的距離学習として学習途中に TNs を変更する簡単なアプローチを提案した. ベンチマークデータに対する数値実験により提案アルゴリズムが状況によって有効に働くことを検証した.

参考文献

[1] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[2] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global framework for nonlinear dimensionallity reduction. *Science*, 290:2319–2323, 2000.

表 2: テストデータに対する最近傍分類の分類誤差の平均と標準偏差. $k = 5$ における実験結果が記されている. ユークリッド距離はテストデータのユークリッド距離における最近傍分類誤差である. 各データの最良の結果になったものを太文字で表す.

	ユークリッド距離		TNs の変更無し		提案アルゴリズム	
	平均 [%]	(± 標準偏差)	平均 [%]	(± 標準偏差)	平均 [%]	(± 標準偏差)
BreastCancerDiagnostic	5.50136	(± 1.15012)	5.17615	(± 1.16001)	**4.68835**	(± 1.18507)
MAGICGammaTelescope	25.6918	(± 1.45652)	**24.1658**	(± 0.99408)	24.5813	(± 1.61106)
Parkinsons	13.5385	(± 5.65500)	13.6923	(± 5.54463)	**12.6154**	(± 4.39951)
SPECTFHeart	24.3820	(± 4.17231)	21.3483	(± 3.51343)	**20.6741**	(± 4.37419)

[3] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems 14*. 2009.

[4] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov. Neighbourhood components analysis. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 513–520. MIT Press, Cambridge, MA, 2005.

[5] Kilian Weinberger, John Blitzer, and Lawrence Saul. Distance metric learning for large margin nearest neighbor classification. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 1473–1480. MIT Press, Cambridge, MA, 2006.

[6] J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon. Information-theoretic metric learning. In *Proceedings of the 24th International Conference on Machine Learning*, pages 209–216, 2007.

[7] T. Hastie and R. Tibshirani. Discriminant adaptive nearest neighbor classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:607–615, 1996.

[8] X. He and P. Niyogi. Locality preserving projections. In S. Thrun, L. Saul, and B. Schökopf, editors, *Advances in Neural Information Processing Systems*, volume 16, 2004.

[9] *IBM ILOG CPLEX 12.0.*