

サポートベクトルにおけるベイズ境界性最大化学習法の検証

野口 大樹[†] 片桐 滋[†] 大崎 美穂[†]
[†] 同志社大学

1. はじめに

パターン認識の本質的な目標であるベイズ誤り(最小分類誤り)状態に対応するベイズ境界を直接的に目指す手法として、ベイズ境界性最大化(MBB: Maximum Bayes Boundary-ness)学習法[1]が近年提案された。MBB 学習法では、推定境界近傍に存在する標本が学習に大きく関与する。そのため境界決定に多大な影響を及ぼす標本のみを学習標本として、MBB 学習法を検証することによりベイズ誤り状態の達成を目指す。本稿では、学習標本としてサポートベクトルを採用する。

2. ベイズ境界性最大化学習法とサポートベクトル

MBB 学習法は境界近傍標本におけるベイズ境界性尺度値(ベイズ境界との類似度)の推定と、勾配法(本稿では最急降下法)による損失の最小化を通じた分類器パラメータの更新を繰り返すことでベイズ境界の達成を目指す。

ベイズ境界性尺度値の推定では推定境界上に関する分類事後確率が求められるが、必ずしも推定境界上に学習標本が存在するとは限らない。そのため、境界近傍標本を用いて分類事後確率の推定を行う。

境界近傍標本の選択について、学習標本の推定境界との距離尺度においてある閾値以内に存在する標本を境界上標本と仮定し、その標本において k 近傍法を用いることで境界近傍標本を抽出する。ここで、本稿での距離尺度は、式(1)に示す誤分類尺度を用いる。

$$d_y(\mathbf{x}; \mathbf{\Lambda}) = -g_y(\mathbf{x}; \mathbf{\Lambda}) + g_{y^*}(\mathbf{x}; \mathbf{\Lambda}), \quad (1)$$

ただし、 $d_y(\mathbf{x}; \mathbf{\Lambda})$ は誤分類尺度値、 \mathbf{x} は入力標本、 $\mathbf{\Lambda}$ は分類器パラメータ、 $g_y(\mathbf{x}; \mathbf{\Lambda})$ は識別関数値、 y^* は入力標本が持つクラスラベルを除いた中で、最も識別関数値が大きくなるクラスラベル(best incorrect class)を表す。誤分類尺度値が正なら誤分類、負なら正分類に対応する。

サポートベクトルを学習標本とするために、学習前にサポートベクトルマシン(SVM: Support Vector Machine)によるフィルター処理を行う。図1にその流れを示す。

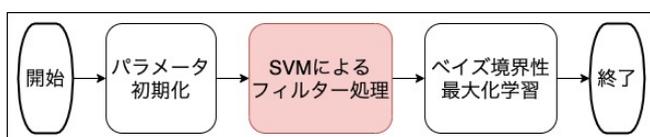


図1. SVM によるフィルター処理を組み込んだ MBB 学習法のフローチャート。

3. 評価実験

本稿では、サポートベクトルを学習標本とした場合の分類誤り率をベイズ誤り率の参考値や先行研究の誤り率と比較することにより評価する。しかし、無限個の標本上で定義されるベイズ誤り率は真値を知ることが不可能であるため、SVM と k 平均法による MPT 型分類器で Cross Validation (CV)を行う。その中の分類誤り率の最小値をベイズ誤り率の参考値としている。CV 法におけるフォールド数は、ほぼ正確なベイズ誤り推定が可能であることを確かめた10とした。図2に結果を示す。なお、CV 法における参考値(紫、青)は学習エポックとは独立して求めているため、図2中で一定の値を示している。

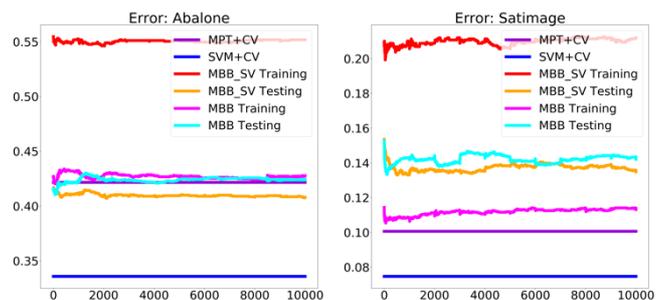


図2. “Abalone”(左)と“Satimage”(右)の実験結果。横軸は学習エポック数、縦軸は分類誤り率を表す。

先行研究(MBB)の Testing(水色)に比べて、本研究(MBB_SV)の Testing(橙)の誤り率が小さく、CV 法におけるベイズ誤り率の参考値(紫、青)に近づいている。つまり、学習目標であるベイズ境界により近づいていることが示唆される。

4. まとめ

本稿では、境界決定に影響を及ぼす標本としてサポートベクトルを採用し、MBB 学習を進めた。その結果、ベイズ境界により近づいている推定境界が得られた。しかしサポートベクトルに限らず、その他の境界決定に影響を及ぼす標本に関しての更なる検証が必要である。

本稿では分類器として MPT 型分類器を採用したが、SVM を分類器として採用した MBB 学習法に関する研究余地についても示唆された。

謝辞: 本研究の一部は科研費(18H03266)の支援を受けた。

参考文献

[1] 千田将大, 他: 信学技報 PRMU2019-101, 2020年3月。