Perceptronの収束性定理と学習率の関係

Yoshihiko Suhara

2012-06-05

概要

パーセプトロンの収束性定理 (Novikoff の定理) [1] の証明を通じて収束性定理が学習率に依存していないことを確認し，パーセプトロンの収束性の保証には学習率が不要なことを確認する．

基本的に文献 [2] を眺めながら自分の言葉でまとめただけなので，[2] をお持ちであればそちらをご覧ください．

1 Perceptron

See Wikipedia.

1.1 アルゴリズム

See Wikipedia.

2 Perceptronの収束性定理

2.1 バイアス項がない場合

定理 1 S を自明でない訓練データ集合とし，$R = \max_{1 \leq i \leq N} \|\mathbf{x}_i\|$ とする．$1 \leq 1 \leq N$ に対して $\|\mathbf{w}^*\| = 1$，かつ $y_i(\langle \mathbf{w}^*, \mathbf{x}_i \rangle) \geq \gamma$ となるようなベクトル $\mathbf{w}^*$ が存在すると仮定する．すると訓練データ集合 S に対し，パーセプトロンアルゴリズムが誤分類する数は高々

$$\frac{R^2}{\gamma^2} \tag{1}$$

となる．

証明

パーセプトロンのアルゴリズムでは重みベクトルは,

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \eta y_i \mathbf{x}_i$$

によって更新されるため,

$$\langle \mathbf{w}^*, \mathbf{w}_t \rangle = \langle \mathbf{w}^*, \mathbf{w}_{t-1} \rangle + \eta y_i \langle \mathbf{x}_i, \mathbf{w}^* \rangle \geq \langle \mathbf{w}^*, \mathbf{w}_{t-1} \rangle + \eta\gamma$$

を得る．ここで

$$\langle \mathbf{w}^*, \mathbf{w}_{t-1} \rangle + \eta\gamma = \langle \mathbf{w}^*, \mathbf{w}_{t-2} \rangle + \eta\gamma + \eta\gamma = \cdots = \langle \mathbf{w}^*, \mathbf{w}_0 \rangle + t\eta\gamma$$

また，$\mathbf{w}_0 = \mathbf{0}$ より，

$$\langle \mathbf{w}^*, \mathbf{w}_t \rangle \geq t\eta\gamma \tag{2}$$

といえる．

また，

$$\begin{aligned}\|\mathbf{w}_t\|^2 &= \|\mathbf{w}_{t-1}\|^2 + 2\eta y_i \langle \mathbf{w}_{t-1}, \mathbf{x}_i \rangle + \eta^2 \|\mathbf{x}_i\|^2 \\ &\leq \|\mathbf{w}_{t-1}\|^2 + \eta^2 \|\mathbf{x}_i\|^2 \\ &\leq \|\mathbf{w}_{t-1}\|^2 + \eta^2 R^2 \end{aligned} \tag{3}$$

ここで $y_i \langle \mathbf{w}_{t-1}, \mathbf{x}_i \rangle \leq 0, \mathbf{x}_i \leq R\ \forall i$ を利用した．ここで

$$\|\mathbf{w}_{t-1}\|^2 + \eta^2 R^2 \leq \|\mathbf{w}_{t-2}\|^2 + \eta^2 R^2 + \eta^2 R^2 \leq \cdots \leq \|\mathbf{w}_0\|^2 + t\eta^2 R^2$$

であるため，式 (3) より，

$$\|\mathbf{w}_t\|^2 \leq t\eta^2 R^2 \tag{4}$$

といえる．

式 (2) の両辺を二乗し，

$$(\langle \mathbf{w}^*, \mathbf{w}_t \rangle)^2 \geq t^2 \eta^2 \gamma^2 \tag{5}$$

を得る[*1]．

また式 (4) の両辺に $\|\mathbf{w}^*\|^2$ をかけて

$$\|\mathbf{w}^*\|^2 \|\mathbf{w}_t\|^2 \leq \|\mathbf{w}^*\|^2 t\eta^2 R^2 \tag{6}$$

を得る．コーシー・シュワルツの不等式より，

$$\|\mathbf{w}^*\| \|\mathbf{w}_t\| \geq \langle \mathbf{w}^*, \mathbf{w}_t \rangle$$

すなわち，

$$\|\mathbf{w}^*\|^2 \|\mathbf{w}_t\|^2 \geq (\langle \mathbf{w}^*, \mathbf{w}_t \rangle)^2$$

のため，式 (5) と式 (6) を用いて，

$$\|\mathbf{w}^*\|^2 t\eta^2 R^2 \geq \|\mathbf{w}^*\|^2 \|\mathbf{w}_t\|^2 \geq (\langle \mathbf{w}^*, \mathbf{w}_t \rangle)^2 \geq t^2 \eta^2 \gamma^2$$

ここで $\|\mathbf{w}^*\| = 1$ より，

$$\begin{aligned} t\eta^2 R^2 &\geq t^2 \eta^2 \gamma^2 \\ \frac{R^2}{\gamma^2} &\geq t \end{aligned}$$

よって定理 1 で示したように，高々 $\frac{R^2}{\gamma^2}$ の試行で収束することが示された．ここで学習率 η が消えており，定理 1 の結果は学習率 η に依存しないことがわかる．したがって，パーセプトロンのアルゴリズムは学習率 η の値によらず収束することが保証される．

[*1] ここで不等式が成り立つために $\eta > 0$ が必要．

2.2 バイアス項がある場合

See [2] (pp.17-20)

(略) パーセプトロンアルゴリズムが誤分類する数は高々

$$\frac{4R^2}{\gamma^2}$$

となる.

参考文献

[1] A. B. Novikoff, "On convergence proofs on perceptrons", Symposiumon the Mathematical Theory of Automata, vol.12, pp.615-622, 1962.

[2] N. Cristianini, J. Shawe-Taylor (大北剛 訳), "サポートベクターマシン入門", 共立出版, 2005.