chapter 5 性能解析

2021/5/3

パーセプトロンの学習定理

(復習)

- 入力ベクトル$\mathsf{x} \in \mathbb{R}^m$およびパラメータ$\mathsf{w} \in \mathbb{R}^m$に対するパーセプトロンの出力

$$\mathrm{sign}\left(\mathsf{w}^T \mathsf{x}\right) \tag{1}$$

- パラメータの初期値: $\mathsf{w}^{(0)} = 0$
- 学習データ$\mathsf{x} \in \mathbb{R}^m, y \in \{-1, 1\}$を受け取り、$\mathrm{sign}\left(\mathsf{w}^T \mathsf{x}\right) \neq y$ならば、次の更新則に従いパラメータを更新する

$$\mathsf{w}^{(t+1)} = \mathsf{w}^{(t)} + y\mathsf{x} \tag{2}$$

パーセプトロンの学習定理

- 性能解析に関連する定義

定義 (学習データのマージン)

学習データ$\{\mathrm{x}^{(t)}, y^{(t)}\}_{t=1,\dots,N}$について、$y^{(t)}\mathrm{u}^T\mathrm{x}^{(t)} \geq \gamma$を満たし$\|\mathrm{u}\| = 1$であるベクトルuが存在するとき、学習データはマージンγで線形分離可能であるという。

定義 (学習データの半径)

学習データ$\{\mathrm{x}^{(t)}, y^{(t)}\}_{t=1,\dots,N}$の半径を$R = \max_t \|\mathrm{x}^{(t)}\|$とする。

パーセプトロンの学習定理

定理 (パーセプトロンの学習定理)

学習データについて、半径がRでありマージンγで分離可能であるならば、この学習データに対するパーセプトロンの更新回数は高々$(R/\gamma)^2$回である。

パーセプトロンの学習定理

Proof.

k回目の更新を考える。

$$\begin{aligned}\mathsf{w}^{(k+1)}\mathsf{u} &= \mathsf{w}^{(k)}\mathsf{u} + y^{(k)}\mathsf{u}^T\mathsf{x}^{(k)} \\ &\geq \mathsf{w}^{(k)}\mathsf{u} + \gamma\end{aligned}$$

ここで、$\mathsf{w}^{(1)} = 0$より、
$k = 1$のとき

$$\mathsf{w}^{(2)}\mathsf{u} \geq \gamma$$

$k = 2$のとき

$$\begin{aligned}\mathsf{w}^{(3)}\mathsf{u} &\geq \mathsf{w}^{(1)}\mathsf{u} + y^{(1)}\mathsf{u}^T\mathsf{x}^{(1)} + \gamma \\ &\geq 2\gamma\end{aligned}$$

パーセプトロンの学習定理

Proof.

と再帰的に導くと、

$$\mathsf{w}^{(k+1)}\mathsf{u} \geq k\gamma \tag{3}$$

また、$y^{(k)}\mathsf{w}^{(k)T}\mathsf{x}^{(k)} < 0, y^2\|\mathsf{x}^{(t)}\|^2 \leq R^2$より、

$$\begin{aligned}\|\mathsf{w}^{(k+1)}\|^2 &= \|\mathsf{w}^{(k)} + y^{(k)}\mathsf{x}^{(k)}\|^2 \\ &\leq \|\mathsf{w}^{(k)}\|^2 + R^2\end{aligned}$$

ここで、先ほどと同様再帰的に導くと、

$$\|\mathsf{w}^{(k+1)}\|^2 \leq kR^2 \tag{4}$$

□

パーセプトロンの学習定理

Proof.

$(3), (4)$およびコーシー・シュワルツの定理より、

$$\begin{aligned} kR^2 &\geq \|\mathsf{w}^{(k+1)}\|^2\|\mathsf{u}\|^2 \\ &\geq \|\mathsf{w}^{(k+1)}\mathsf{u}\|^2 \\ &\geq k^2\gamma^2 \end{aligned} \tag{5}$$

(5)より、

$$k \leq (R/\gamma)^2 \tag{6}$$

□

パーセプトロンの学習

- パーセプトロンの更新回数の鄭玄は学習データの半径とマージンの大きさにのみ依存する
- よって、高次元データであってもスパースであれば更新回数を抑えることができる

線形分離可能でない場合のパーセプトロンの学習定理

ref. Freund and Schapire (1999)

定義 (学習データのペナルティ)

$\|\mathrm{u}\| = 1$であるベクトルuと$\gamma > 0$が
与えられたとき、学習データのペナルティを次のように定義する。

$$d^{(t)} = \max\{0, \gamma - y^{(t)}\mathrm{u}^T\mathrm{x}^{(t)}\}$$

定義 (学習データのペナルティのノルム)

ペナルティノルムを次のように定義する。

$$D = \sqrt{\Sigma_{i=1}^{n} d^{(t)2}} \tag{7}$$

線形分離可能でない場合のパーセプトロンの学習定理

定理 (線形分離可能でない場合のパーセプトロンの学習定理)

学習データの半径をR、$\gamma > 0$のときのペナルティノルムをDとする。このとき、パーセプトロンの更新回数は高々$(R + D/\gamma)^2$回である。