感知机

★ 收敛性

定理 2.1（Novikoff） 设训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$ 是线性可分的，其中 $x_i\in\mathcal{X}=\mathbf{R}^n$， $y_i\in\mathcal{Y}=\{-1,+1\}$， $i=1,2,\cdots,N$， 则

（1）存在满足条件 $\|\hat{w}_{\mathrm{opt}}\|=1$ 的超平面 $\hat{w}_{\mathrm{opt}}\cdot\hat{x}=w_{\mathrm{opt}}\cdot x+b_{\mathrm{opt}}=0$ 将训练数据集完全正确分开；且存在 $\gamma>0$， 对所有 $i=1,2,\cdots,N$

$$y_i(\hat{w}_{\mathrm{opt}}\cdot\hat{x}_i)=y_i(w_{\mathrm{opt}}\cdot x_i+b_{\mathrm{opt}})\geqslant\gamma \tag{2.8}$$

（2）令 $R=\max\limits_{1\leqslant i\leqslant N}\|\hat{x}_i\|$， 则感知机算法 2.1 在训练数据集上的误分类次数 k 满足不等式

$$k\leqslant\left(\frac{R}{\gamma}\right)^2 \tag{2.9}$$

★ 原始模式

算法 2.1（感知机学习算法的原始形式）

输入：训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$，其中 $x_i\in\mathcal{X}=\mathbf{R}^n$， $y_i\in\mathcal{Y}=\{-1,+1\}$， $i=1,2,\cdots,N$； 学习率 $\eta\,(0<\eta\leqslant1)$；

输出： w,b； 感知机模型 $f(x)=\mathrm{sign}(w\cdot x+b)$.

（1）选取初值 w_0,b_0

（2）在训练集中选取数据 (x_i,y_i)

（3）如果 $y_i(w\cdot x_i+b)\leqslant0$

$$w\leftarrow w+\eta y_i x_i$$
$$b\leftarrow b+\eta y_i$$

（4）转至（2），直至训练集中没有误分类点. ■

这种学习算法直观上有如下解释：当一个实例点被误分类，即位于分离超平面的错误一侧时，则调整 w, b 的值，使分离超平面向该误分类点的一侧移动，以减少该误分类点与超平面间的距离，直至超平面越过该误分类点使其被正确分类.

算法 2.1 是感知机学习的基本算法，对应于后面的对偶形式，称为原始形式．感知机学习算法简单且易于实现.

★ 对偶形式

算法 2.2（感知机学习算法的对偶形式）

输入：线性可分的数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$，其中 $x_i\in\mathbf{R}^n$， $y_i\in\{-1,+1\}$， $i=1,2,\cdots,N$； 学习率 η（$0<\eta\leqslant1$）；

输出： α,b； 感知机模型 $f(x)=\mathrm{sign}\left(\sum\limits_{j=1}^{N}\alpha_j y_j x_j\cdot x+b\right)$.

其中 $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_N)^{\mathrm{T}}$.

（1） $\alpha\leftarrow0$， $b\leftarrow0$

（2）在训练集中选取数据 (x_i,y_i)

（3）如果 $y_i\left(\sum\limits_{j=1}^{N}\alpha_j y_j x_j\cdot x_i+b\right)\leqslant0$

$$\alpha_i\leftarrow\alpha_i+\eta$$
$$b\leftarrow b+\eta y_i$$

（4）转至（2）直到没有误分类数据. ■

对偶形式中训练实例仅以内积的形式出现．为了方便，可以预先将训练集中实例间的内积计算出来并以矩阵的形式存储，这个矩阵就是所谓的 Gram 矩阵（Gram matrix）

$$G=[x_i\cdot x_j]_{N\times N}$$