Lecture2 作业

1，假设训练样本集为$D = \{(\mathbf{x}_1, y_1) = ((3,3)^T, 1), (\mathbf{x}_2, y_2) = ((4,3)^T, 1), (\mathbf{x}_3, y_3) = ((1,1)^T, -1)\}$, 使用感知器算法设计分类面，并判断测试样本$\mathbf{x} = (0,1)^T$属于哪个类别。

解：

样本增广后为：$\vec{x}_1 = (1,3,3)^T$，$y_1 = 1$，$\vec{x}_2 = (1,4,3)^T$，$y_2 = 1$，$\vec{x}_3 = (1,1,1)^T$，$y_3 = -1$

初始化权重：$\vec{w}^{(0)} = (0,0,0)^T$

$sign\left(\vec{w}^{(0)T}\vec{x}_1\right) = 0 \neq y_1$， $\therefore$ $\vec{w}^{(1)} = \vec{w}^{(0)} + y_1\vec{x}_1 = (1,3,3)^T$，

$sign\left(\vec{w}^{(1)T}\vec{x}_2\right) = 1 = y_2$， $\therefore$ $\vec{w}^{(2)} = \vec{w}^{(1)} = (1,3,3)^T$

$sign\left(\vec{w}^{(2)T}\vec{x}_3\right) = 1 \neq y_3$， $\therefore$ $\vec{w}^{(3)} = \vec{w}^{(2)} + y_3\vec{x}_3 = (0,2,2)^T$

$sign\left(\vec{w}^{(3)T}\vec{x}_1\right) = 1 = y_1$， $\therefore$ $\vec{w}^{(4)} = \vec{w}^{(3)} = (0,2,2)^T$

$sign\left(\vec{w}^{(4)T}\vec{x}_2\right) = 1 = y_2$， $\therefore$ $\vec{w}^{(5)} = \vec{w}^{(4)} = (0,2,2)^T$

$sign\left(\vec{w}^{(5)T}\vec{x}_3\right) = 1 \neq y_3$， $\therefore$ $\vec{w}^{(6)} = \vec{w}^{(5)} + y_3\vec{x}_3 = (-1,1,1)^T$

$sign\left(\vec{w}^{(6)T}\vec{x}_1\right) = 1 = y_1$， $\therefore$ $\vec{w}^{(7)} = \vec{w}^{(6)} = (-1,1,1)^T$

$sign\left(\vec{w}^{(7)T}\vec{x}_2\right) = 1 = y_2$， $\therefore$ $\vec{w}^{(8)} = \vec{w}^{(7)} = (-1,1,1)^T$

$sign\left(\vec{w}^{(8)T}\vec{x}_3\right) = 1 \neq y_3$， $\therefore$ $\vec{w}^{(9)} = \vec{w}^{(8)} + y_3\vec{x}_3 = (-2,0,0)^T$

$sign\left(\vec{w}^{(9)T}\vec{x}_1\right) = -1 \neq y_1$， $\therefore$ $\vec{w}^{(10)} = \vec{w}^{(9)} + y_1\vec{x}_1 = (-1,3,3)^T$

$sign\left(\vec{w}^{(10)T}\vec{x}_2\right) = 1 = y_2$， $\therefore$ $\vec{w}^{(11)} = \vec{w}^{(10)} = (-1,3,3)^T$

$sign\left(\vec{w}^{(11)T}\vec{x}_3\right) = 1 \neq y_3$， $\therefore$ $\vec{w}^{(12)} = \vec{w}^{(11)} + y_3\vec{x}_3 = (-2,2,2)^T$

$sign\left(\vec{w}^{(12)T}\vec{x}_1\right) = 1 = y_1$， $\therefore$ $\vec{w}^{(13)} = \vec{w}^{(12)} = (-2,2,2)^T$

$sign\left(\vec{w}^{(13)T}\vec{x}_2\right) = 1 = y_2$， $\therefore$ $\vec{w}^{(14)} = \vec{w}^{(13)} = (-2,2,2)^T$

$sign\left(\vec{w}^{(14)T}\vec{x}_3\right) = 1 \neq y_3$， $\therefore$ $\vec{w}^{(15)} = \vec{w}^{(14)} + y_3\vec{x}_3 = (-3,1,1)^T$

$sign\left(\vec{w}^{(15)T}\vec{x}_1\right) = 1 = y_1$， $\therefore$ $\vec{w}^{(16)} = \vec{w}^{(15)} = (-3,1,1)^T$

$sign\left(\vec{w}^{(16)T}\vec{x}_2\right) = 1 = y_2$， $\therefore$ $\vec{w}^{(17)} = \vec{w}^{(16)} = (-3,1,1)^T$

$sign\left(\vec{w}^{(17)T}\vec{x}_3\right) = -1 = y_3$， $\therefore$ $\vec{w}^{(18)} = \vec{w}^{(17)} = (-3,1,1)^T$

$\therefore$ $\vec{w} = (-3,1,1)^T$, 分类面为：$x_1 + x_2 - 3 = 0$

对测试样本进行增广，$\vec{\mathrm{x}} = (1,0,1)^T$，

$$sign\left(\vec{w}^T\vec{\mathrm{x}}\right) = sign((-3,1,1)(1,0,1)^T = -1, \quad \therefore \quad \vec{x} \in -1 \text{ 类}$$

2，对于感知器算法（PLA），假设第 t 次迭代时，选择的是第 n 个样本：$\mathrm{sign}(\boldsymbol{w}^T\mathbf{x}_n) \neq y_n$，$\boldsymbol{w}_{t+1} \leftarrow \boldsymbol{w}_t + y_n\mathbf{x}_n$，下述那个式子正确？

（a）$\boldsymbol{w}_{t+1}^{\boldsymbol{T}}\mathbf{x}_n = y_n$

（b）$\mathrm{sign}\left(\boldsymbol{w}_{t+1}^{\boldsymbol{T}}\mathbf{x}_n\right) = y_n$

（c）$y_n\boldsymbol{w}_{t+1}^{\boldsymbol{T}}\mathbf{x}_n \geq y_n\boldsymbol{w}_t^{\boldsymbol{T}}\mathbf{x}_n$

（d）$y_n\boldsymbol{w}_{t+1}^{\boldsymbol{T}}\mathbf{x}_n < y_n\boldsymbol{w}_t^{\boldsymbol{T}}\mathbf{x}_n$

3，证明：针对线性可分训练样本集，PLA 算法中，当$\mathbf{W}_0 = \mathbf{0}$，在对分错样本进行了 T 次纠正后，下式成立：$\frac{\mathbf{w}_f^T}{\|\mathrm{w}_f\|}\frac{\mathbf{w}_T}{\|\mathbf{w}_T\|} \geq \sqrt{T} \cdot constant$

证明：由于

$$\begin{aligned} \boldsymbol{W}_f^T \boldsymbol{W}_{t+1} &= \boldsymbol{W}_f^T (\boldsymbol{W}_t + y_n(t)\boldsymbol{X}_n(t)) \\ &\geqslant \boldsymbol{W}_f^T \boldsymbol{W}_t + \min_n y_n(t) \boldsymbol{W}_f^T \boldsymbol{X}_n(t) \end{aligned}$$

且有$W_0=0$，故有$\boldsymbol{W}_f^T \boldsymbol{W}_T \geqslant T \cdot \min_n y_n \boldsymbol{W}_f^T \boldsymbol{X_n}$；

又由于

$$\begin{aligned} \|\boldsymbol{W}_{t+1}\|^2 &= \|\boldsymbol{W}_t + y_n(t)\boldsymbol{X}_n(t)\|^2 \\ &= \|\boldsymbol{W}_t\|^2 + 2y_n(t)\boldsymbol{W}_t^T \boldsymbol{X}_n(t) + \|y_n(t)\boldsymbol{X}_n(t)\|^2 \\ &\leqslant \|\boldsymbol{W}_t\|^2 + 0 + \|y_n(t)\boldsymbol{X}_n(t)\|^2 \\ &\leqslant \|\boldsymbol{W}_t\|^2 + \max_n \|\boldsymbol{X}_n(t)\|^2 \end{aligned}$$

故有$\|\boldsymbol{W}_T\| \leqslant \sqrt{T \cdot \max_n \|\boldsymbol{X}_n\|^2}$；

综上所述，有

$$\begin{aligned} \frac{\boldsymbol{W}_f^T \boldsymbol{W}_T}{\|\boldsymbol{W}_f\|\|\boldsymbol{W}_T\|} &\geqslant \frac{T \cdot \min_n y_n \boldsymbol{W}_f^T \boldsymbol{X_n}}{\|\boldsymbol{W}_f\| \cdot \sqrt{T \cdot \max_n \|\boldsymbol{X}_n\|^2}} \\ &= \sqrt{T} \cdot constant \end{aligned}$$

4，针对线性可分训练样本集，PLA 算法中，假设对分错样本进行了 T 次纠正后得到的分类面不再出现错分状况，定义：$\boldsymbol{R}^2 = \max_n \|\boldsymbol{x}_n\|^2$ ，$\rho = \min_n y_n \frac{\mathbf{w}_f^T}{\|\mathbf{w}_f\|}\mathbf{x}_n$，试证明：$\mathrm{T} \leq \frac{\boldsymbol{R}^2}{\rho^2}$

证明：

$$\begin{aligned} \frac{\boldsymbol{W}_f^T \boldsymbol{W}_T}{\|\boldsymbol{W}_f\|\|\boldsymbol{W}_T\|} &\geqslant \frac{T \cdot \min_n y_n \boldsymbol{W}_f^T \boldsymbol{X_n}}{\|\boldsymbol{W}_f\| \cdot \sqrt{T \cdot \max_n \|\boldsymbol{X}_n\|^2}} \\ &= \sqrt{T} \cdot \frac{\rho}{R} \end{aligned}$$

$$
\begin{aligned}
\sqrt{T} &\leqslant \frac{R}{\rho} \cdot \frac{\boldsymbol{W}_f^T \boldsymbol{W}_T}{\|\boldsymbol{W}_f\| \|\boldsymbol{W}_T\|} \\
&= \frac{R}{\rho} \cdot \cos\langle \boldsymbol{W}_f, \boldsymbol{W}_T \rangle \\
&\leqslant \frac{R}{\rho}
\end{aligned}
$$

因此有

$$
T \leqslant \frac{R^2}{\rho^2}
$$

5， 假设训练样本集为 $D=\{(\vec{x}_1,y_1)=((0.2,0.7)^T,1),(\vec{x}_2,y_2)=((0.3,0.3)^T,1),(\vec{x}_3,y_3)=((0.4,0.5)^T,1)$，$(\vec{x}_4,y_4)=((0.6,0.5)^T,1)$，$(\vec{x}_5,y_5)=((0.1,0.4)^T,1)$，$(\vec{x}_6,y_6)=((0.4,0.6)^T,-1)$，$(\vec{x}_7,y_7)=((0.6,0.2)^T,-1)$，$(\vec{x}_8,y_8)=((0.7,0.4)^T,-1)$，$(\vec{x}_9,y_9)=((0.8,0.6)^T,-1)$，$(\vec{x}_{10},y_{10})=((0.7,0.5)^T,-1)\}$，用 Pocket 算法设计分类面。（可借助编程实现，迭代次数最多 10 次，需提交每次迭代的结果）

解：略