THU-70250043, Pattern Recognition (Spring 2018) **Homework: 4**

Linear Classifiers and SVM

Lecturer: Changshui Zhang `zcs@mail.tsinghua.edu.cn`

Student: XXX `xxx@mails.tsinghua.edu.cn`

Problem 1

Suppose we have N points x_i in $\mathcal{R}^p$ in general position, with class labels $y_i \in \{-1, 1\}$. Prove that the perceptron learning algorithm converges to a separating hyperplane in a finite number of steps:

- Denoting a hyperplane by $f(x) = \beta_1^T(x) + \beta_0 = 0$, or in more compact notation $\beta^T x^* = 0$, where $x^* = (x, 1)$ and $\beta = (\beta_1, \beta_0)$. Let $z_i = \frac{x_i^*}{||x_i^*||}$. Show that separability implies the existence of a β_{sep} such that $y_i \beta_{sep}^T z_i \geq 1, \forall i$
- Given a current β_{old}, the perceptron algorithm identifies a point z_i that is misclassified, and produces the update $\beta_{new} = \beta_{old} + y_i z_i$. Show that $||\beta_{new} - \beta_{sep}||^2 \leq ||\beta_{old} - \beta_{sep}||^2 - 1$, and hence the algorithm converges to a separating hyperplane in no more than $||\beta_{start} - \beta_{sep}||^2$ steps. (Ripley, 1996)

Problem 2

Consider a dataset with 2 points in 1d: $(x_1 = 0, y_1 = -1)$and $(x_2 = \sqrt{2}, y_2 = 1)$. Consider mapping each point to 3d using the feature vector $\phi = [1, \sqrt{2}x, x^2]^T$. (This is equivalent to using a second order polynomial kernel.) The max margin classifier has the form

$$min||w||^2 \quad s.t. \tag{1}$$

$$y_1(w^T\phi(x_1) + w_0) \quad \geq \quad 1 \tag{2}$$

$$y_2(w^T\phi(x_2) + w_0) \quad \geq \quad 1 \tag{3}$$

a. Write down a vector that is parallel to the optimal vector w.

b. What is the value of the margin that is achieved by this w? Hint: recall that the margin is the distance from each support vector to the decision boundary. Hint 2: think about the geometry of 2 points in space, with a line separating one from the other.

c. Solve for w, using the fact the margin is equal to $1/||w||$.

d. Write down the form of the discriminant function$f(x) = w_0 + w^T \phi(x)$ as an explicit function of x.

Program

经典感知器（Algorithm 1）的训练过程可以看成在解区内寻找一个解，并没有对这个解的性能有所限定。这个解只需满足 $\alpha^T y_n > 0$，其中 α 是感知器的权向量，y_n 是规范化增广样本向量。而 margin 感知器（Algorithm 2）则要求算法收敛的超平面有一个大于 γ 的 margin，其中 γ 是预先设定的一个正数。即， margin 感知器的解需要满足 $\alpha^T y_n > \gamma$。

Algorithm 1 Fixed-Increment Single Sample Correction Algorithm

initialize α,$k \leftarrow 0$
repeat
    $k \leftarrow (k+1)$ mod n
    if y_k is misclassified by α **then**
        $\alpha \leftarrow \alpha + y_k$
    end if
until all patterns are properly classified
return α

因此，在 margin 感知器的训练中，“错误”包括两种情况：1）标签预测错误（prediction mistake）；2）标签预测正确但 margin 不够大（margin mistake）。

Algorithm 2 Single Sample Correction Algorithm With Margin

initialize $\alpha, k \leftarrow 0,\ \gamma > 0$
repeat
    $k \leftarrow (k+1)$ mod n
    if $\alpha^T y_k \leq \gamma$ **then**
        $\alpha \leftarrow \alpha + y_k$
    end if
until all patterns are properly classified with a large enough margin γ
return α

- 随机生成 200 个二维平面上的点，其中 100 个标记为 1，剩下的 100 个标记为 -1，保证它们是线性可分的。画出这 200 个点的分布。

- 编程实现经典感知器算法，并在生成的数据集上运行。在一张图上画出分界线和数据点。

- 编程实现 margin 感知器算法，并在生成的数据集上运行。在一张图上画出分界线和数据点，分析 γ 取值对算法收敛性及分界面位置的影响。