ST790 HW6

Steven Xu

November 5, 2019

1.

(a)

Since the data are linearly separable, there exists $\boldsymbol{\beta}$ such that

$$y_i \boldsymbol{\beta}^T \mathbf{x}_i > 0,\ \forall i = 1, ..., N$$
$$y_i \frac{\boldsymbol{\beta}^T \mathbf{x}_i}{||\mathbf{x}_i||} = y_i \boldsymbol{\beta}^T \mathbf{x}_i^* > 0,\ \forall i = 1, ..., N$$

further we have

$$y_i \boldsymbol{\beta}^T \mathbf{x}_i^* \geq \min_i \left(y_i \boldsymbol{\beta}^T \mathbf{x}_i^* \right),\ \forall i = 1, ..., N$$

Let $m = \min_i \left(y_i \boldsymbol{\beta}^T \mathbf{x}_i^* \right)$, then

$$\because\ m > 0$$
$$\therefore\ \frac{1}{m} y_i \boldsymbol{\beta}^T \mathbf{x}_i^* \geq 1,\ \forall i = 1, ..., N$$

It suffice to let $\boldsymbol{\beta}_{sep} = \frac{1}{m}\boldsymbol{\beta}$.

(b)

$$\begin{aligned}
||\boldsymbol{\beta}_{new} - \boldsymbol{\beta}_{sep}||^2 &= \boldsymbol{\beta}_{new}^T \boldsymbol{\beta}_{new} - 2\boldsymbol{\beta}_{new}^T \boldsymbol{\beta}_{sep} + \boldsymbol{\beta}_{sep}^T \boldsymbol{\beta}_{sep} \\
||\boldsymbol{\beta}_{old} - \boldsymbol{\beta}_{sep}||^2 &= \boldsymbol{\beta}_{old}^T \boldsymbol{\beta}_{old} - 2\boldsymbol{\beta}_{old}^T \boldsymbol{\beta}_{sep} + \boldsymbol{\beta}_{sep}^T \boldsymbol{\beta}_{sep} \\
\therefore\ ||\boldsymbol{\beta}_{new} - \boldsymbol{\beta}_{sep}||^2 - ||\boldsymbol{\beta}_{old} - \boldsymbol{\beta}_{sep}||^2 &= \boldsymbol{\beta}_{new}^T \boldsymbol{\beta}_{new} - 2\boldsymbol{\beta}_{new}^T \boldsymbol{\beta}_{sep} - \boldsymbol{\beta}_{old}^T \boldsymbol{\beta}_{old} + 2\boldsymbol{\beta}_{old}^T \boldsymbol{\beta}_{sep} \\
&= \left(\boldsymbol{\beta}_{old} + y_i \mathbf{z}_i\right)^T \left(\boldsymbol{\beta}_{old} + y_i \mathbf{z}_i\right) - 2\left(\boldsymbol{\beta}_{old} + y_i \mathbf{z}_i\right)^T \boldsymbol{\beta}_{sep} - \boldsymbol{\beta}_{old}^T \boldsymbol{\beta}_{old} + 2\boldsymbol{\beta}_{old}^T \boldsymbol{\beta}_{sep} \\
&= 2 y_i \boldsymbol{\beta}_{old}^T \mathbf{z}_i + ||y_i \mathbf{z}_i||^2 - 2 y_i \boldsymbol{\beta}_{sep}^T \mathbf{z}_i
\end{aligned}$$

Using the fact that $\mathbf{z}_i$ is misclassified given $\boldsymbol{\beta}_{old}$, we have

$$y_i \boldsymbol{\beta}_{old}^T \mathbf{z}_i < 0$$
$$\because\ y_i \boldsymbol{\beta}_{sep} \mathbf{z}_i \geq 1,\ |y_i| = 1,\ ||\mathbf{z}_i||^2 = 1$$
$$\therefore 2 y_i \boldsymbol{\beta}_{old}^T \mathbf{z}_i + ||y_i \mathbf{z}_i||^2 - 2 y_i \boldsymbol{\beta}_{sep}^T \mathbf{z}_i \leq -2 + 1 = -1$$

Suppose $\boldsymbol{\beta}_n = \boldsymbol{\beta}_{sep}$

$$\begin{aligned}
||\boldsymbol{\beta}_n - \boldsymbol{\beta}_{sep}||^2 &\leq ||\boldsymbol{\beta}_{n-1} - \boldsymbol{\beta}_{sep}||^2 - 1 \\
||\boldsymbol{\beta}_n - \boldsymbol{\beta}_{sep}||^2 &\leq ||\boldsymbol{\beta}_{n-2} - \boldsymbol{\beta}_{sep}||^2 - 2 \\
&\cdots \\
||\boldsymbol{\beta}_n - \boldsymbol{\beta}_{sep}||^2 &\leq ||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{sep}||^2 - n \\
\therefore\ n &\leq ||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{sep}||^2
\end{aligned}$$

where $\boldsymbol{\beta}_1 := \boldsymbol{\beta}_{start}$.

2.

Here I list the best combination of tuning parameters for each kernel selection.

Table 1: Scenario 2

kernel	cost	gamma	degree	coef0	Testing Error
Linear	0.146	N/A	N/A	N/A	0.211
Polynomial	0.125	0.445	5	1.2	0.177
Gaussian	0.197	3.190	N/A	N/A	0.176

Based on these results, the Gaussian kernel with the listed parameter choice should be used as the final model.

3.

Here I list the best combination of tuning parameters for each kernel selection.

Table 2: ZIP data

kernel	cost	gamma	degree	coef0	Testing Error
Linear	0.006	N/A	N/A	N/A	0.027
Polynomial	0.016	0.112	3	1.667	0.016
Gaussian	0.645	0.002	N/A	N/A	0.022

Based on these results, the polynomial kernel with the listed parameter choice should be used as the final model.