TANMAY GUPTA
tg289

MATH-680

HW-5

(1)

Test errors ->

For Linear SVM = 0.0467

For Gaussian Kernel SVM = 0.027

For polynomial Kernel SVM = 0.030

Gaussian Kernel SVM performs the best with a test error of 0.027 as compared to 0.0467 for ~~Kernel~~ Linear SVM & 0.030 for polynomial kernel SVM.
[shown in R code]

(2)

By using deviance in splits of the tree, we got

Training error = 0.048

Test error = 0.093

(3)

(a)

If the points are separable then there exists a vector β such that =>

$$\beta^T x_i^* > 0 \quad \text{when } y_i = +1$$

$$\beta^T x_i^* < 0 \quad \text{when } y_i = -1$$

for all $i = 1, 2, \ldots N$, so

$$y_i \beta^T x_i^* \geq 0 \quad \text{for all } i$$

$$y_i \beta^T > 0 \quad (\text{dividing by } \|x_i^*\|)$$

let $m > 0$ be the smallest value of $y_i \beta^T z_i$ over all training set, so

$$y_i \beta^T z_i \geq m \quad \text{for all } i$$

we divide both sides by m

$$y_i \left(\frac{1}{m} \beta\right)^T z_i \geq 1$$

As $\beta_{sep} = \frac{1}{m} \beta$

So $\underline{y_i \beta^T_{sep} z_i \geq 1}$ for all i ($\forall i$)

(b) $\beta_{new} = \beta_{old} + y_i z_i$

$\Rightarrow \beta_{new} - \beta_{sep} = \beta_{old} - \beta_{sep} + y_i z_i$

$$\|\beta_{new} - \beta_{sep}\|^2 = \|\beta_{old} - \beta_{sep}\|^2 + \underline{y_i^2 \|z_i\|^2} + 2y_i(\beta_{old} - \beta_{sep})^T z_i$$

As $y_i = \pm 1$ & $\|z_i\|^2 = 1$, so $\underline{y_i^2 \|z_i\|^2 = 1}$

Now, as $y_i z_i$ was misclassified by β_{old}

so, $\underline{y_i \beta^T_{old} z_i \leq 0}$

β_{sep} can correctly classify all points

so, $\underline{y_i \beta^T_{sep} z_i \geq 1}$

$$2y_i(y_i \beta^T_{old} z_i - y_i \beta^T_{sep} z_i) \leq 2(0-1) = -2$$

$$\|\beta_{new} - \beta_{sep}\|^2 \leq \|\beta_{old} - \beta_{sep}\|^2 + 1 - 2$$

$$\leq \underline{\|\beta_{old} - \beta_{sep}\|^2 - 1}$$

So, we can drive any initial vector β_{start} to β_{sep} in at most $\|\beta_{start} - \beta_{sep}\|^2$ steps.