SAHIL JINDAL

2013CS10253

GROUP 5

# COL774: Assignment2

## 1.

**(a)** For linear kernel,

SVM dual objective:

$$\max\left(\sum_{i=1}^{m}\alpha_i - \frac{1}{2} * \sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}\left(x^{(i)^T}x^{(j)}\right)\right)$$

Constraints:

$$\sum_{i=0}^{m}\alpha_i y^{(i)} = 0 \text{ , and } 0 \leq \alpha_i \leq C$$

Converting dual objective to form: $\alpha^T Q\alpha + b^T\alpha + c$, we get

$$Q = -\frac{1}{2} * ((Y * Y^T).* (X * X^T))$$

$b =$ Column vector consisting of ones.

Total support vectors obtained = 281

**(b)** Weight vector $w = \sum_{i=0}^{m}\alpha_i y^{(i)}x^{(i)}$

Intercept term $b = -\frac{1}{2}(\max_{i:y^{(i)}=-1} w^T x^{(i)} + \min_{i:y^{(i)}=1} w^T x^{(i)})$

Average Accuracy = 98.33%

**(c)** For Gaussian kernel,

SVM dual objective:

$$\max\left(\sum_{i=1}^{m}\alpha_i - \frac{1}{2} * \sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}e^{\left(-\left\|x^{(i)}-x^{(j)}\right\|^2 * \gamma\right)}\right)$$

$$Q = -\frac{1}{2} * ((Y * Y^T).* K$$

$$K_{ij} = e^{\left(-\|x^{(i)}-x^{(j)}\|^2 * \gamma\right)}$$

$$\gamma = 2.5 * 10^{-4}$$

Accuracy = 98.72%

Total support vectors = 435

Accuracy obtained in case of Gaussian kernel is higher as compared to that for linear kernel.

**(d)** Using LibSVM library,

Support vectors obtained in case of Linear kernel = 319

Support vectors obtained in case of Gaussian kernel = 646

Accuracy obtained in case of Linear kernel = 98.20%

Accuracy obtained in case of Gaussian kernel = 98.72%

Comparing results of CVX package and LibSVM:

- Results obtained using LibSVM are significantly faster than time taken by CVX package.
- Accuracy obtained in both the cases is almost similar.
- Different algorithms are implemented in the packages.

## 2.

**(b)** $\theta$'s for hidden and input layers are initialized randomly between -0.025 and 0.025.

Number of units in output layer = 1

The output from the unit present in output layer is > 0.5 => digit predicted is 3, and output < 0.5 => digit predicted = 8.

Stopping criteria:

$$|\Delta J(\theta)| < 10^{-5}$$

**(c)** Total iterations executed = 28

Time taken per iteration = 748/28 = 26.71 sec.

Average accuracy obtained over test3 and test8 = 98.89%

**(d)** Output units required in this case = 10

For digit $= i$, $(i+1)^{th}$ output unit should produce output > 0.5 and all other output units should produce output < 0.5.

Stopping Criteria:

$$|\Delta J(\theta)| < 10^{-4}$$

Total iterations executed = 32

Time taken per iteration = 4565/32 = 142.65 sec.

Average Accuracy obtained over test data = 93.46%

As compared to binary setting, the time taken in training has increased significantly. This is because in each iteration we have more sample entries in this case as compared to binary setting.