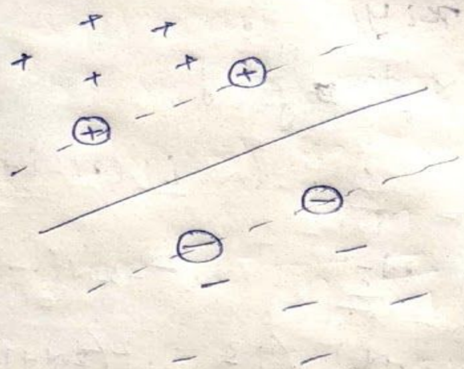
CSE512 Fall 2018 - Machine Learning - Homework 4

Name: Manideep Attanti
Solar Id: 112028167
Netid email: manideep.attanti@stonybrook.edu

Q1.1 and 1.2

Q 1.1



Circled points are support vectors.

SVMs main objective is to ~~minimize~~ maximize the margin i.e maximize the minimum distance of the points from the classifier plane.

$$\text{LOOCV error} = \sum_{i=1}^{n} f(y^i, x_i : \theta^{-i}\ b^{-i})$$

$$y^i \text{ expected} = x_i^T \theta^{-i} + b^{-i}$$

where θ^{-i} is the value of weights leaving i^{th} example & b^{-i} is the bias leaving i^{th} example.

if $y^i \times y^i$expected $= 1$ (correctly classified)

else (misclassified)

and value of $f(x) = 0$ if $x = 1$, else 1

Now considering 2 cases:

(i) If a point is not a support vector. Then the margins donot change since the margins are solely defined by support vectors

$\therefore f(x) = 0$

(ii) If a point is a support vector. Then the margins will shift giving a new classifier plane. This may lead to the case where this point is misclassified.

$\therefore f(x) = 0$ if correctly classified

$= 1$ if misclassified.

In the worst case, it is possible that all the support vectors may be mis classified

$\therefore$ no. of support vectors is m

$$\text{LOOCV error} = \frac{1}{n} \sum_{i=1}^{m} f(y^i, x^i ; \theta^{-i}, b^{-i})$$

where x^i is a support vector for whole data.

For worst case scenario $f(x) = 1$ $\forall x \in$ support vectors

$$\therefore \text{max LOOCV error} = \frac{1}{n} \sum_{i=1}^{m} (1)$$

$$= \frac{m}{n}$$

$$\therefore \text{LOOCV error} \leq \frac{m}{n}$$

Q1.2 The data is linearly seperable in higher dimension. Even in this case, since the point are linearly seperable, the margin is decided by support vector. Hence the loss function f in 1.1 will evaluate to 0 for non- support vectors and may evaluate to 1 for a support vector. Hence above concept holds. and LOOCV $\leq m/n$

Q2.1

Q2.1) $\underset{\alpha}{\text{maximize}} \sum_{j=1}^{n} \alpha_j - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i \alpha_i y_j \alpha_j \cdot k(x_i, x_j)$

$$\text{s.t.} \sum_{j=1}^{n} y_j \alpha_j = 0$$

$$0 \leq \alpha_j \leq C \quad \forall j$$

which is same as

$\underset{\alpha}{\text{minimize}} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i \alpha_i y_j \alpha_j \cdot k(x_i, x_j) - \sum_{j=1}^{n} \alpha_j$

$$\text{s.t} \sum_{j=1}^{n} y_j \alpha_j = 0$$

$$0 \leq \alpha_j \leq C \quad \forall j$$

Quadprog in matlab:

$$\min_{x} \frac{1}{2} x^T H x + f^T x \quad \text{s.t.} \begin{cases} A \cdot x \leq b \\ Aeq \cdot x = beq \\ lb \leq x \leq ub \end{cases}$$

Since we are using linear kernel,

$H = (Y Y^T) .* (X^T X)$ $\quad .* \rightarrow$ element wise product

$f = -1 \times \text{ones}(\text{size}(Y,1),1) \equiv \begin{bmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{bmatrix}_{n \times 1}$

$\left.\begin{array}{l} A = [\ ] \\ b = [\ ] \end{array}\right\} \rightarrow$ empty matrices

$Aeq = \begin{bmatrix} y_1, & y_2, & \cdots, & y_n \\ 0, & 0, & \cdots, & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & & 0 \end{bmatrix}_{n \times n}$ first row is Y^T and rest all elements are zeroes

$beq = \text{zeros}(\text{size}(Y,1),1) \equiv \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{n \times 1}$

$lb = \text{zeros}(\text{size}(Y,1),1) \equiv \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{n \times 1}$

$ub = C \times \text{ones}(\text{size}(Y,1),1) \equiv \begin{bmatrix} C \\ \vdots \\ C \end{bmatrix}_{n \times 1}$

Q2.4)
For C = 0.1

Accuracy:	90.76%
Objective Function:	24.765
Number of Support Vectors:	339

Confusion Matrix:

152	2
32	181

Q2.5)
For C = 10

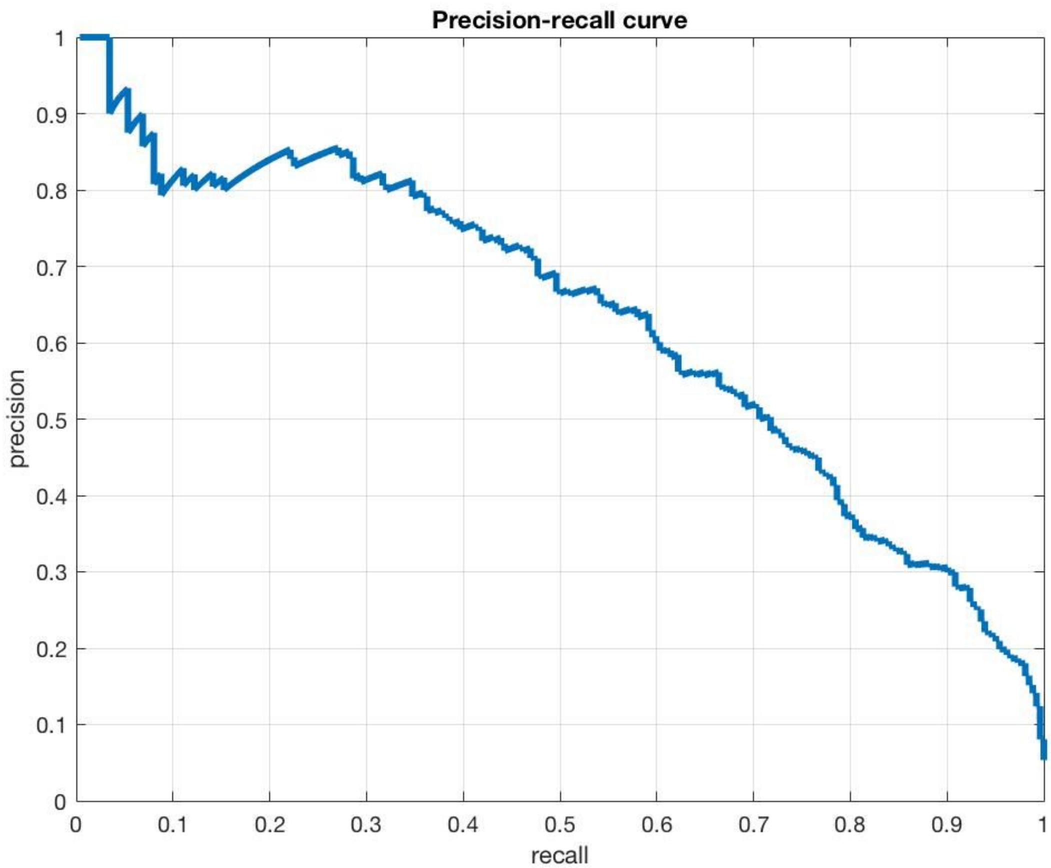
Accuracy:	97.82%
Objective Function:	112.146
Number of Support Vectors:	123

Confusion Matrix:

180	4
4	179

Q2.6)
Best accuracy achieved is 0.78369 based on Kaggle submission.
Used One Vs All approach to solve for multi-class. Used Linear Kernel and C value is 0.1.
Changing the C value had no impact on accuracy.

Q3.4.1)
AP: 0.636



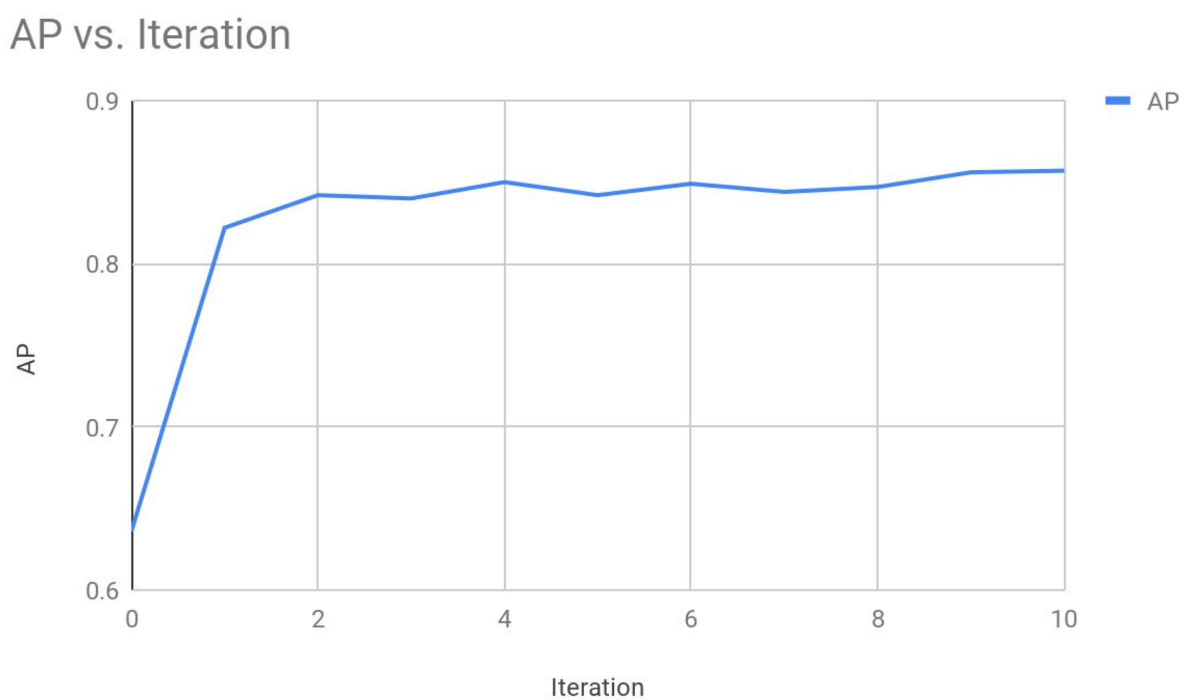
Q3.4.3)
Objective Values:

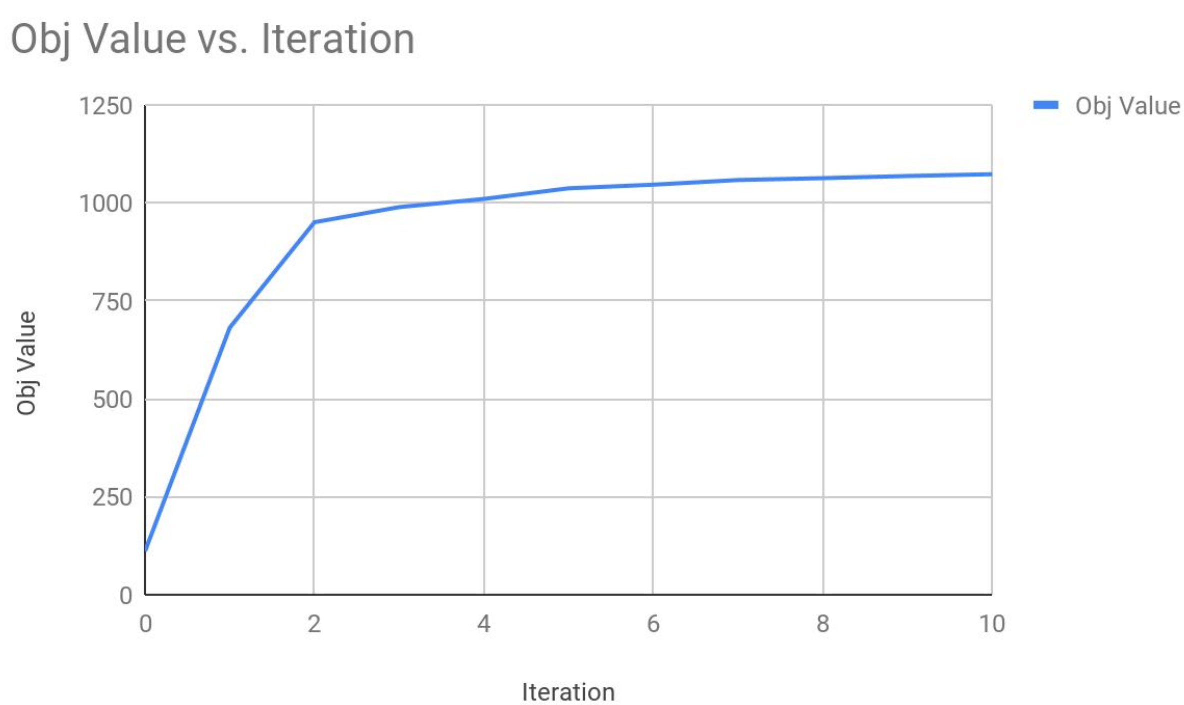
Iteration	0	1	2	3	4	5	6	7	8	9	10
Obj Value	111.171	681.34	950.48	989.03	1009.90	1037.04	1046.28	1058.20	1062.92	1068.46	1072.97

AP Values:

Iteration	0	1	2	3	4	5	6	7	8	9	10
AP	0.636	0.822	0.842	0.840	0.850	0.842	0.849	0.844	0.847	0.856	0.857

0th index is the values before starting hard negative mining.
C = 10





Q3.4.4
AP: 0.813638