1.1

$$L = (\bar{x}^T w - y)^2 + \lambda \|w\|^2$$

derivating both sides & putting $\frac{\partial L}{\partial w} = 0$

$$\frac{\partial L}{\partial w} = 0 = 2\bar{x}(\bar{x}^T w - y) + 2\lambda \bar{w}$$

$$2\bar{x}\cdot\bar{x}^T\bar{w} - 2\bar{x}y + 2\lambda\bar{w} = 0$$

putting $d = \bar{x}y$ & $c = \bar{x}\cdot\bar{x}^T + \lambda I$

$$(\bar{x}\bar{x}^T + \lambda \hat{I})\bar{w} = \bar{x}y$$

$$c\cdot\bar{w} = d \qquad \boxed{\bar{w} = c^{-1}d}$$

1.2

$$C_{(i)} = [x_{i-1}; x_{i+1}]^* [x_{i-1}; x_{i+1}]^T + \lambda I$$

$$C_{(i)} = \bar{x}\cdot x_i^T - x_i\cdot x_i^T + \lambda I$$

$$\boxed{C(i) = C - x_i x_i^T}$$ putting $\bar{x}_i x^T + \lambda I = c$

$$d = \bar{x}y$$

$$d_i = \bar{x}_i y_i$$

$$d(i) = [d_{i-1}; d_{i+1}] = [x_{i-1}; x_{i+1}]\cdot[y_{i-1}; y_{i+1}]$$

$$= \bar{x}\cdot y - \bar{x}_i\cdot y_i$$

$$\boxed{d(i) = d - \bar{x}_i y_i}$$

1.3 From previous question,

$$C_{(i)} = C - x_i x_i^T$$

$$C_{(i)}^{-1} = \left(C - x_i x_i^T\right)^{-1}$$

$$= C^{-1} + \frac{C^{-1}\bar{x}_i \bar{x}_i^T C^{-1}}{1 + \bar{x}_i^T C^{-1}(-\bar{x}_i)}$$

[using Sherman-Morrison formula & putting values $A = C$, $u = -\bar{x}_i$, $V^T = \bar{x}_i^T$]

$$= C^{-1} + \frac{C^{-1}\bar{x}_i \bar{x}_i^T C^{-1}}{1 - \bar{x}_i^T C^{-1}\bar{x}_i}$$

1.4

$$\overline{w_{(i)}} = C_{(i)}^{-1} d_{(i)}$$

$$= \left(C^{-1} + \frac{C^{-1}\bar{x}_i \bar{x}_i^T C^{-1}}{1 - \bar{x}_i^T C^{-1}\bar{x}_i}\right)(d - \bar{x}_i y_i)$$

$$= C^{-1}d - C^{-1}\bar{x}_i y_i + \frac{C^{-1}\bar{x}_i \bar{x}_i^T C^{-1} d - C^{-1}\bar{x}_i \bar{x}_i^T C^{-1}\bar{x}_i y_i}{1 - \bar{x}_i^T C^{-1}\bar{x}_i}$$

$$= \bar{w} + C^{-1}\bar{x}_i \left[\frac{-y_i(1 - \bar{x}_i^T C^{-1}\bar{x}_i) + \bar{x}_i^T C^{-1} d - \bar{x}_i^T C^{-1}\bar{x}_i y_i}{1 - \bar{x}_i^T C^{-1}\bar{x}_i}\right]$$

$$= \bar{w} + C^{-1}\bar{x}_i \left[\frac{-y_i + C^{-1}\bar{w}}{1 - \bar{x}_i^T C^{-1}\bar{x}_i}\right] \quad \text{[putting } \bar{w} = C^{-1}d\text{]}$$

1.5

$$\bar{w}_{(i)} = \bar{w} + (C^{-1}x_i)\frac{-y_i + \bar{x_i}^T\bar{w}}{1 - \bar{x}^T C^{-1}\bar{x_i}}$$

$\cancel{\bar{x_i}}$

$$\bar{w}_{(i)}^T = \bar{w}^T + \bar{x_i}^T C^{-1T}\frac{(w^T x_i - y_i^T)}{1 - \bar{x_i}^T C^{-1}\bar{x_i}}$$

multiply $\bar{x_i}$ both side,

$$\bar{w}_{(i)}^T \cdot \bar{x}_i = \bar{w}^T\bar{x}_i + \bar{x}_i \cdot \bar{x_i}^T C^{-1T} \cdot \frac{w^T \bar{x}_i - y_i^T}{1 - \bar{x_i}^T C^{-1}\bar{x_i}}$$

$$= \frac{\bar{w}^T\bar{x_i} - x_i\bar{x_i}^T C^{-1}\bar{w}^T x_i + \bar{x}_i\bar{x}_i^T C^{-1T}\overline{w^T\bar{x}_i - y_i^T} \cdot \bar{x}_i}{1 - \bar{x_i}^T C^{-1}\bar{x}_i}$$

$$\cancel{\bar{w}_{(i)}^T\bar{x}_i - y_i = \frac{\bar{w}^T\bar{x}_i - \bar{y}_i^T(}{1 - \bar{x}_i^T C^{-1}\bar{x}_i} - y_i}$$

$$\bar{w}_{(i)}^T\bar{x_i} - y_i = \frac{\bar{w}^T\bar{x_i} - \bar{x}_i^T C^{-1}\bar{x_i}\cdot\bar{y}_i - (1 - x_i^T C^{-1}\bar{x_i})\cdot y_i}{1 - x_i^T C^{-1}\bar{x_i}}$$

$$= \frac{\bar{w}^T\bar{x_i} - y_i}{1 - \bar{x_i}^T C^{-1}\bar{x}_i}$$

1.6

$$LOOCV = \sum_{i=1}^{n} (\bar{\omega}_{(i)}^{T} \bar{x}_i - y_i)^2$$

c^{-1} takes $O(k^3)$ time only once.

~~on~~ all other vector ~~matrix~~ multiplication will take $O(k^2)$ time

This will be executed n times. so ~~to~~tal complexity will be $O(nk^2)$

Total complexicity $= O(k^3 + nk^2)$

Q-2

2.1 Here,

x_1 = boolean variable

x_2 = continuous variable

y = boolean variable.

$$y_{pred} = \underset{y_k}{\operatorname{argmax}} \frac{P(Y=y_k)\,\Pi_i P(x_i \mid Y=y_k)}{\sum_j P(Y=y_j)\,\Pi_i P(x_i \mid Y=y_j)}$$

$$= \underset{y_k}{\operatorname{argmax}} \frac{P(Y=y_k) \cdot P(x_1 \mid Y=y_k) \cdot P(X_2 \mid Y=y_k)}{\sum_j P(Y=y_j) \cdot P(x_1 / Y=y_j) \cdot P(X_2 / Y=y_j)}$$

param required for predicting y,

for var X_1,

x_1 is boolean variable which follows bernoulli distribution.

$P(X_1 \mid Y=y_k)$ $\Rightarrow$ $\theta_1^1 = P(X_1 = T / Y = T)$ $\quad \theta_1^2 = 1 - \theta_1^1$

$\theta_2^1 = P(X_1 = T / Y = F)$ $\quad \theta_2^2 = 1 - \theta_2^1$

let $\cancel{\theta_1^1 = P(X_1 / Y = True)}$ $\quad \cancel{\theta_1^2 = P(X_1 / Y = False)} = \cancel{1-\theta_1^1}$

$\cancel{\theta_2^1 = P(X_1 = False / Y = True)}$ $\quad \cancel{\theta_2^2 = P(X_1 = False / Y = False)} = \cancel{1-\theta_2^1}$

we can find out θ_1^2 & θ_2^2 using θ_1^1 & θ_2^1

so, we need to estimate only two parameters.

for variable x_2,

x_2 will follow normal distribution. for normal distribution we need to find out two parameters μ & σ. this we need for both True & False class. so, we need to find out 4 params.

$$\mu_1^2 = \frac{1}{\sum \delta(y_j = T)} \sum_j x_2^j \delta(y^j = T) \Rightarrow \text{when } y \text{ is True}$$

$$\mu_2^2 = \frac{1}{\sum \delta(y_j = F)} \sum_j x_2^j \delta(y^j = F) \Rightarrow \text{when } y \text{ is False}$$

$$(\sigma_1^2)^2 = \frac{1}{\sum \delta(y_j = T)} \sum_j (x_2^j - \mu_1^2)^2 \cdot \delta(y^j = T)$$

$$(\sigma_2^2)^2 = \frac{1}{\sum \delta(y_j = F)} \sum_j (x_2^j - \mu_2^2)^2 \cdot \delta(y^j = F)$$

putting all these values in Prev. eqn

$$P(y=T/x) = \frac{1}{1 + \frac{P(y=F) \cdot P(X|y=F)}{P(y=T) \cdot P(X/y=T)}}$$

$$= \frac{1}{1 + \exp\left(\ln \frac{P(y=F)}{P(y=T)} + \ln\left(\frac{P(X|y=F)}{P(X/y=T)}\right)\right)}$$

$$= \frac{1}{1 + \exp\left(\ln\left(\frac{P(y=F)}{P(y=T)}\right) + \ln\left(\frac{P(x_1|y=F)}{P(x_1|y=T)}\right) + \ln\left(\frac{P(x_2|y=F)}{P(x_2|y=T)}\right)\right)}$$

as discussed earlier, x_1 follows bernulli distribution,

$$P(x_1/y=F) = (\theta_1^2)^{x_i} \cdot (1-\theta_1^2)^{(1-x_i)}$$

$$P(x_1/y=T) = (\theta_1^1)^{x_i} \cdot (1-\theta_1^1)^{(1-x_i)}$$

~~for x_2~~

$$P(x_2|y=F) = \frac{1}{\sqrt{2\pi}\,\sigma_2^2} e^{-\frac{1}{2}\left(\frac{x-\mu_2^2}{\sigma_2^2}\right)^2}$$

$$P(x_2/y=T) = \frac{1}{\sqrt{2\pi}\,\sigma_1^2} e^{-\frac{1}{2}\left(\frac{x-\mu_1^2}{\sigma_1^2}\right)^2}$$

putting all these in the eqn.

$$\ln\left(\frac{P(x_2|y=F)}{P(x_2|y=T}\right) = \ln\left(\frac{\sigma_1^2}{\sigma_2^2} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu_2^2}{\sigma_2^2}\right)^2 + \frac{1}{2}\left(\frac{x-\mu_1^2}{\sigma_1^2}\right)^2}\right)$$

As assumed in class, $\sigma^2 = \sigma_2^2 = \sigma_1^2$

$$\ln\left(\frac{P(x_2|y=F)}{P(x_2|y=T)}\right) = \ln\left(\exp\left(\frac{-(x-\mu_2^2)^2 + (x-\mu_1^2)^2}{2\sigma^2}\right)\right)$$

$$= \frac{(\mu_1^2-\mu_2^2)^2 x_2 + (\mu_1^2)^2 + (\mu_2^2)^2}{2\sigma^2}$$

~~for x_1~~

$$\ln\left(\frac{P(x_1|y=F)}{P(x_1|y=T)}\right) = \ln\left(\frac{(\theta_1^2)^{x_1} \cdot (1-\theta_1^2)^{(1-x_1)}}{(\theta_1^1)^{x_1}(1-\theta_1^1)^{(1-x_1)}}\right)$$

$$= x_1 \ln \theta_1^2 + (1-x_1)\ln(1-\theta_1^2) - x_1 \ln\theta_1^1 - (1-x_1)\ln(1-\theta_1^1)$$

$$= x_1 \ln\left(\frac{\theta_1^2}{\theta_1^1}\right) + (1-x_1)\ln\left(\frac{1-\theta_1^2}{1-\theta_1^1}\right)$$

substitue value of this in final equation.

$$P(y=T/x) = \frac{1}{1+\exp\left(\ln\frac{P(y=F)}{P(y=T)} + x_1\ln\left(\frac{\theta_1^2}{\theta_1^1}\right) + (1-x_1)\ln\left(\frac{1-\theta_1^2}{1-\theta_1^1}\right) + \frac{(\mu_1^2-\mu_2^2)^2 x_2 + (\mu_1^2)^2 + (\mu_2^2)^2}{2\sigma^2}\right)}$$

$$
= \frac{1}{1 + \exp\left(\ln\left(\frac{P(Y=F)}{P(Y=T)}\right) + \ln\left(\frac{1-\theta_1^2}{1-\theta_1^1}\right) + \frac{(\mu_1^2)^2 + (\mu_2^2)^2}{2\sigma^2} + X_1\left(\ln\frac{\theta_1}{\theta_2} - \ln\left(\frac{1-\theta_1^2}{1-\theta_1^1}\right)\right) + \frac{X_2}{2}\left(\frac{\mu_1^2 - \mu_2^2}{\sigma}\right)^2 \right.}
$$

2.2.

$(X_1, X_2 \dots X_d)$ is a vector of booleam variables

Y is also booleam.

In class we showed that,

$$P(Y=1/X) = \frac{P(Y=1) \cdot P(X/Y=1)}{P(Y=1) \cdot P(X/Y=1) + P(Y=0)\, P(X/Y=0)}$$

$$= \frac{1}{1 + \frac{P(Y=0) \cdot P(X/Y=0)}{P(Y=1) \cdot P(X/Y=1)}}$$

$$= \frac{1}{1 + \exp\left(\ln\left(\frac{P(Y=0)}{P(Y=1)}\right) + \sum_i \ln \frac{P(X_i/Y=0)}{P(X_i/Y=1)}\right)}$$

$$= \frac{1}{1 + \exp\left(\ln \frac{1-\pi}{\pi} + \sum_{i=0}^{d} \ln \frac{P(X_i/Y=0)}{P(X_i/Y=1)}\right)}$$

where $\pi = P(Y=1)$

& $1-\pi = P(Y=0)$

[here, $P(X_i/Y=0)$ & $P(X_i=1/Y=1)$ follows bernulli distribution

so, $P(X_i/Y=0) = \theta_{i,0}(1-\theta_{i,0})^{(1-X_i)}$

$P(X_i/Y=1) = \theta_{i,1}(1-\theta_{i,1})^{(1-X_i)}$]

$$= \frac{1}{1 + \exp\left(\ln \frac{1-\pi}{\pi} + \sum_{i=0}^{d} \frac{\theta_{i,0}(1-\theta_{i,0})^{(1-X_i)}}{\theta_{i,1}(1-\theta_{i,1})^{(1-X_i)}}\right.}$$

$$= \frac{1}{1+\exp\left(\ln\frac{1-\pi}{\pi} + \sum_{i=0}^{d}\left(\ln\frac{\theta_{i,0}}{\theta_{i,1}}\right)^{x_i} + \ln\left(\frac{1-\theta_{i,0}}{1-\theta_{i,1}}\right)^{(1-x_i)}\right)}$$

$$= \frac{1}{1+\exp\left(\ln\frac{1-\pi}{\pi} + \sum_{i=1}^{d} x_i \ln\left(\frac{\theta_{i,0}}{\theta_{i,1}}\right) + (1-x_i)\ln\left(\frac{1-\theta_{i,0}}{1-\theta_{i,1}}\right)\right)}$$

$$= \frac{1}{1+\exp\left(\ln\frac{1-\pi}{\pi} + \sum_{i=0}^{d} x_i\left(\ln\left(\frac{\theta_{i,0}}{\theta_{i,1}}\right) - \ln\left(\frac{1-\theta_{i,0}}{1-\theta_{i,1}}\right)\right) + \ln\left(\frac{1-\theta_{i,0}}{1-\theta_{i,1}}\right)\right)}$$

$$= \frac{1}{1+\exp\left(\left(\ln\frac{1-\pi}{\pi}\right) + \sum_{i=0}^{d}\ln\left(\frac{1-\theta_{i,0}}{1-\theta_{i,1}}\right)\right) + \sum_{i=0}^{d} x_i\left(\ln\left(\frac{\theta_{i,0}}{\theta_{i,1}}\right) - \ln\left(\frac{1-\theta_{i,0}}{1-\theta_{i,1}}\right)\right)}$$

$$\boxed{p(y=1/x) = \frac{1}{1+\exp(w^T x)}}$$

here $w_0 = \cancel{\exp}\ \ln\left(\frac{1-\pi}{\pi}\right) + \sum_{i=0}^{d}\ln\left(\frac{1-\theta_{i,0}}{1-\theta_{i,1}}\right)$

$w_i = \sum_{i=0}^{d} x_i\left(\ln\left(\frac{\theta_{i,0}}{\theta_{i,1}}\right) - \ln\left(\frac{1-\theta_{i,0}}{1-\theta_{i,1}}\right)\right)$

This equation is in form of logistic regression.

Q-3

3.1.1

$H :- (Y \cdot Y^T) .* (X^T \cdot X)$

where Y is lable & X is input

$f = -\text{ones}(1,N) = \vec{1}_{1\times N} = [1,1 \ldots 1]_N$

$A = \text{zeros}(1,N) = \vec{0}_{1\times N} = [0,0, \ldots 0]_N$

$b = 0$

$Aeq = Y^T$

$beq = 0$

$Lb = \text{zeros}(N,1) = \vec{0}_{N\times 1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_N$

$UB = C * \text{ones}(N,1) = \vec{C}_{N\times 1} = \begin{bmatrix} C \\ C \\ C \\ C \end{bmatrix}_N$

3.1.4

C = 0.1

accuracy = 97.27

Obj. value = 35.06

Support vector = 5

Confusion matrix :-

177	7
3	180

3.1.5

C = 10

accuracy = 82.01

Obj value = 624.5

Support vector = 121

Confusion matrix =

184	0
66	117

3.2

1. $L_i = \frac{1}{2n} \sum_{j=1}^{k} \|w_j\|^2 + cL(w, x_i, y_i)$

derivating with Respect to w_{y_i}

$$\frac{\partial L_i}{\partial w_{y_i}} = \frac{1}{n} w_{y_i} + c \frac{\partial}{\partial w_{y_i}} L(w, x_i, y_i)$$

$$\frac{\partial L(w, x_i, y_i)}{\partial w_{y_i}} \Rightarrow 0 \quad \text{if } w_{\hat{y}_i}^T - w_{y_i}^T x_i + 1 < 0$$

$$-x_i \quad \text{otherwise}$$

$$\frac{\partial L_i}{\partial w_{y_i}} = \frac{1}{n} w_{y_i} - c x_i$$

2. derivating with Respect to $w_{\hat{y}_i}$

$$\frac{\partial L_i}{\partial \hat{y}_i} = \frac{1}{n} w_{\hat{y}_i} + c * \frac{\partial}{\partial w_{\hat{y}_i}} L(w, x_i, y_i)$$

$$\frac{\partial L}{\partial w_{\hat{y}_i}}(w, x_i, y_i) \Rightarrow = 0 \quad \text{if } w_{\hat{y}_i}^T - w_{y_i}^T x_i + 1 <= 0$$

$$= x_i \quad \text{otherwise}$$

$$\frac{\partial L}{\partial w_{\hat{y}_i}} = \frac{1}{n} w_{y_i} + c x_i$$

3. derivating L_i w.R.T w_j

$$\frac{\partial L(w, x_i, y_i)}{\partial w_j} = \frac{1}{2} \sum_{j=1}^{k} w_j$$

3.2.5

loss for $c = 0.1$ $\Rightarrow$ 19.857

loss for $c = 10$ $\Rightarrow$ 80.44

comparing with objective function of 3.1.4
it is performing better.
These values were 35.87 & 624.5 respectively
for 3.1

accuracy = 96.45

3.2.6

	$c = 0.1$	$c = 10$
(a) prediction (test) accuracy:	0.93733	0.9646
prediction (test) error:	0.063	0.0354
(b) training accuracy:	0.9585	0.9945
training error :-	0.041	0.0055

(c) $\sum_{j=1}^{K} \|w_j\|^2$ = for $c = 0.1$ $\Rightarrow$ 15.0741 120.79

$$\cancel{\frac{1}{1+\exp\left(\ln\frac{1-\pi}{\pi}+\ln\frac{1-\theta_{i,0}}{1-\theta_{i,1}}+\right.}}$$

3.2.7

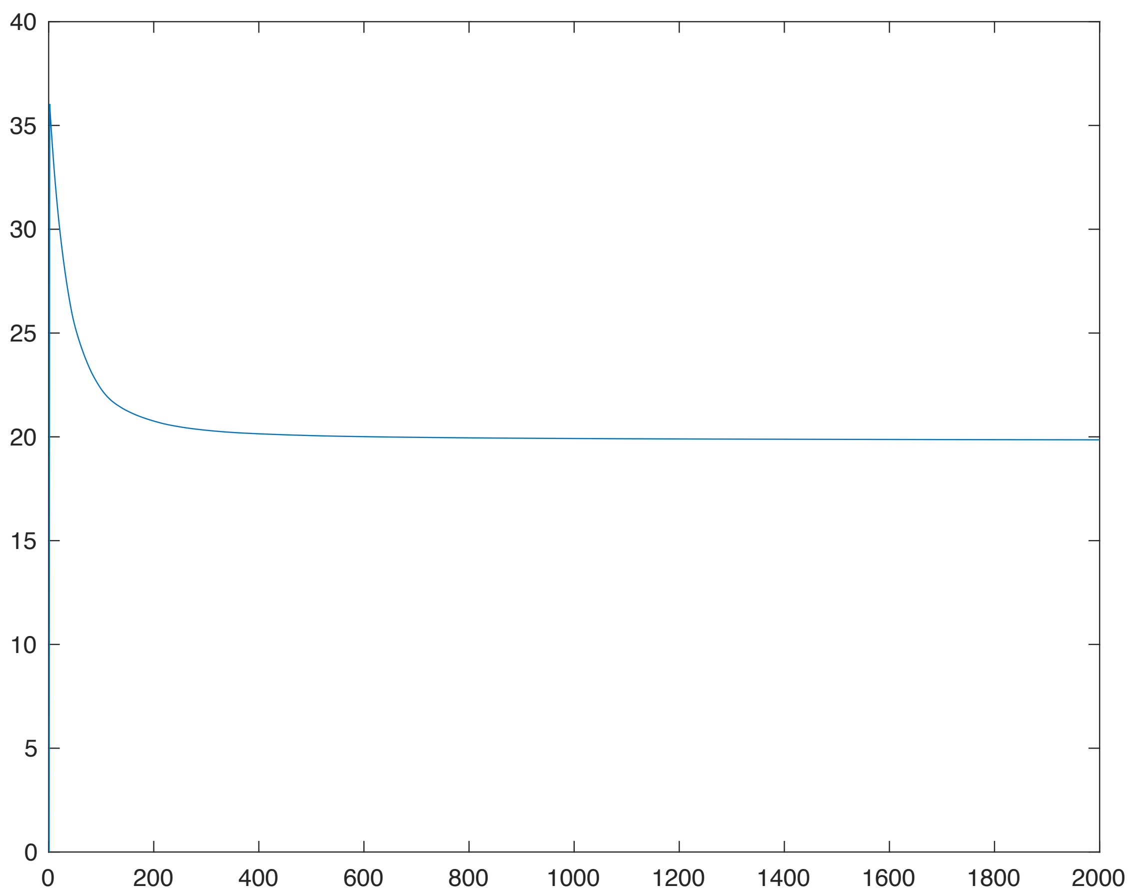
best accuracy :- 79.01 %

Parameters :- $n_0 = 0$

$n_1 = 100$

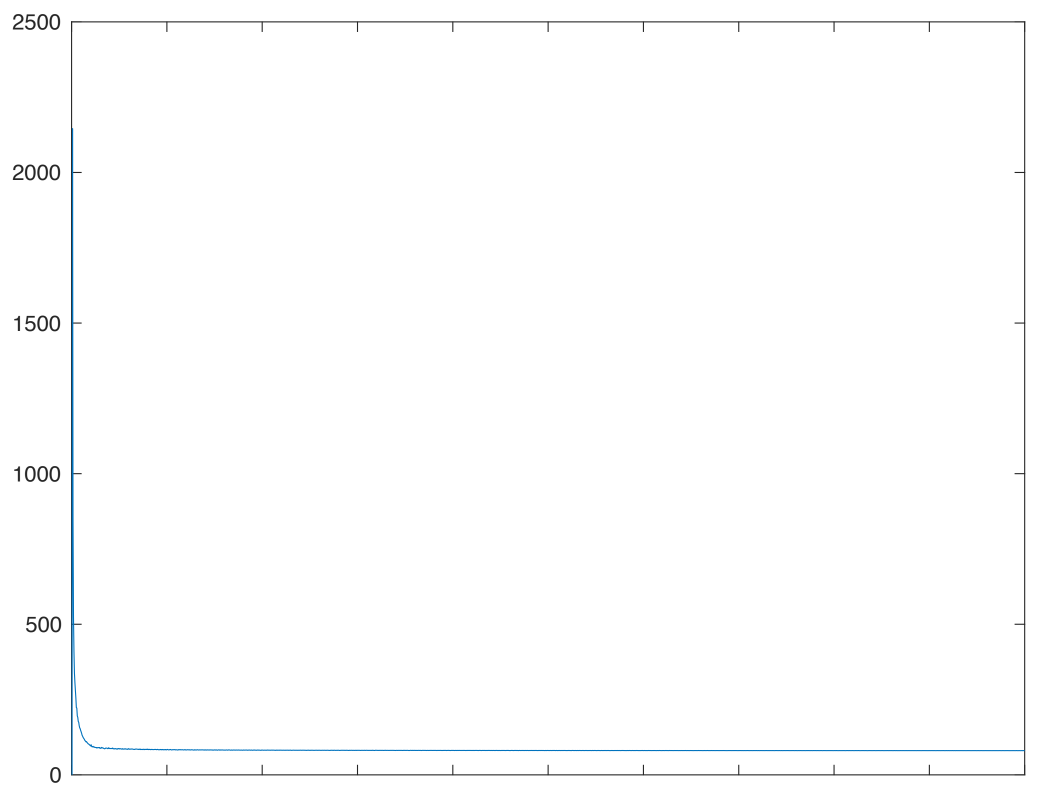
$c = 0.008$

Q 3.2.5 - C = 0.1



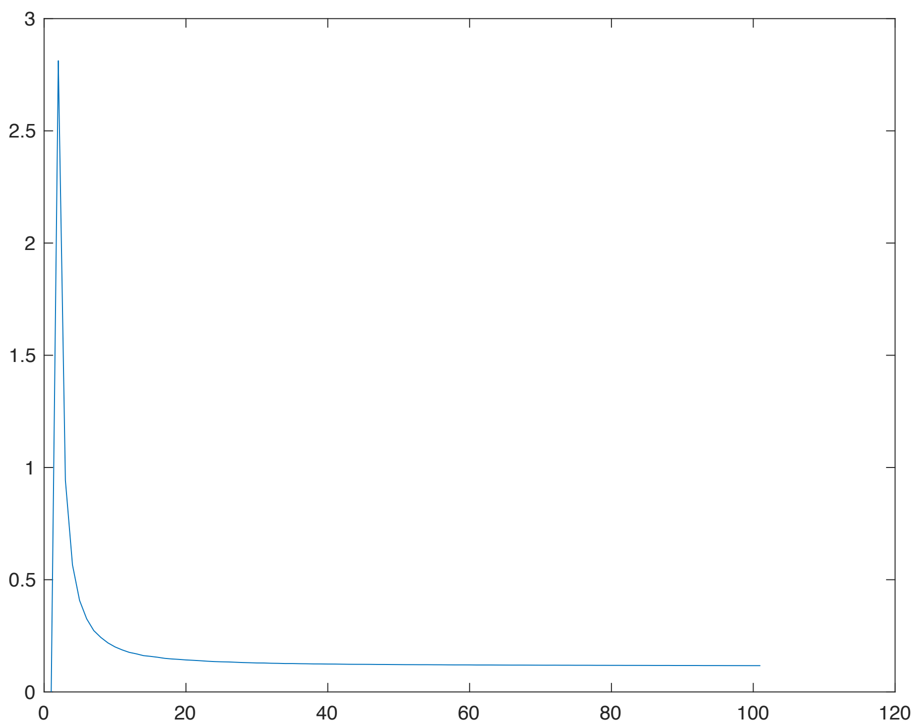
40
35
30
25
20
15
10
5
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

Q 3.2.5 - C = 10



2500
2000
1500
1000
500
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

Q 3.2.7



4.4.1

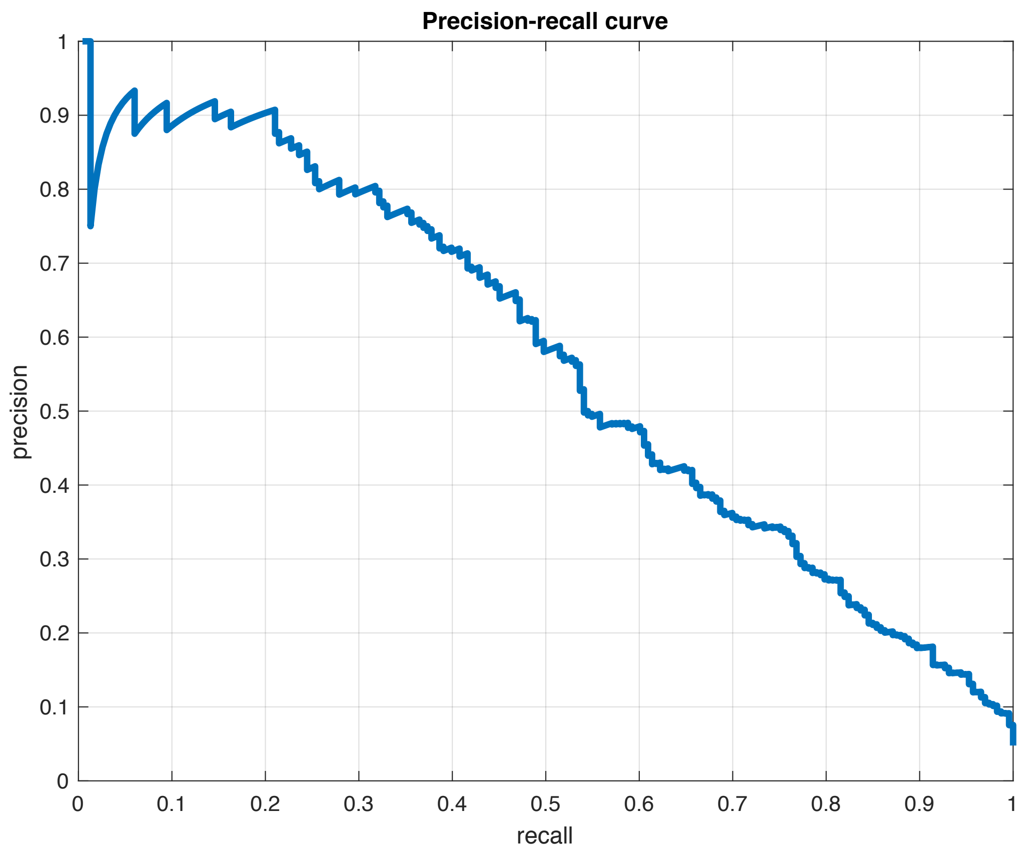
$AP = 0.5732$

4.4.3 plots

4.4.4

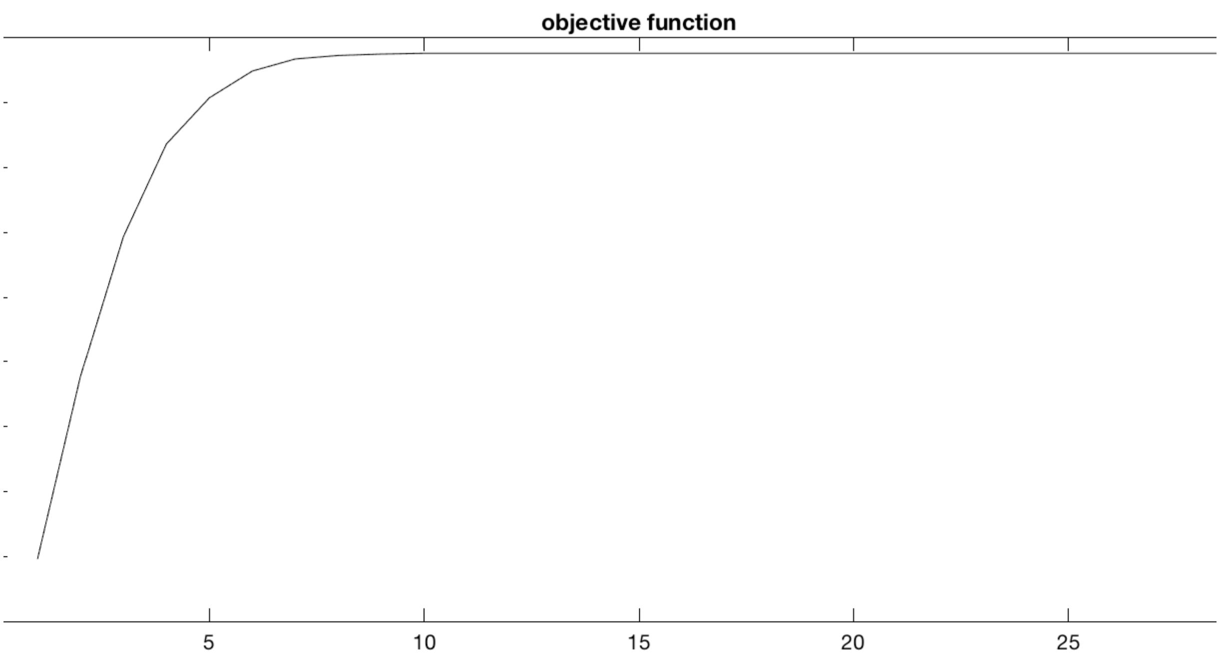
$AP = 0.6$

Q 4.4.1 plot

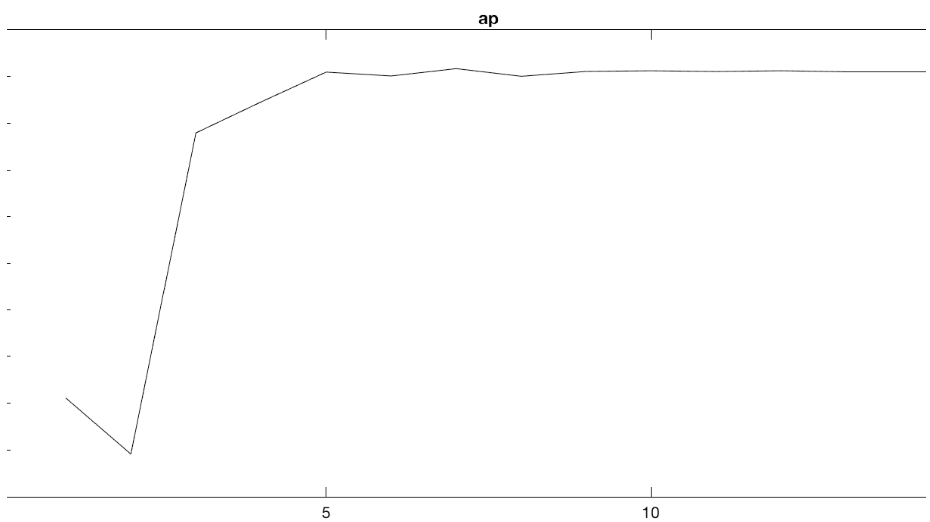


Precision-recall curve
precision
recall
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

4.4.3 Objective Function



4.4.3 APs on validation data



4.4.3 Precision Recall