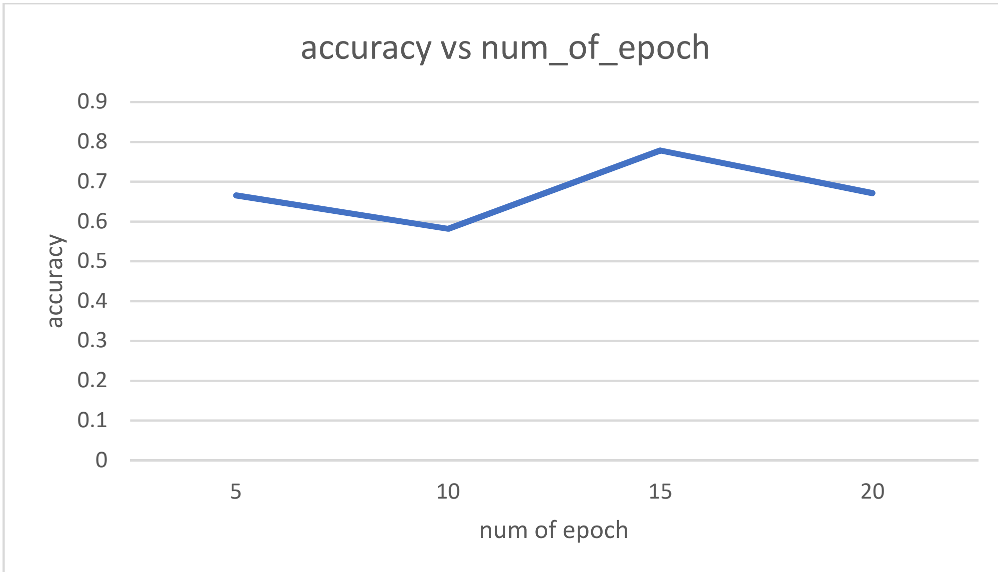
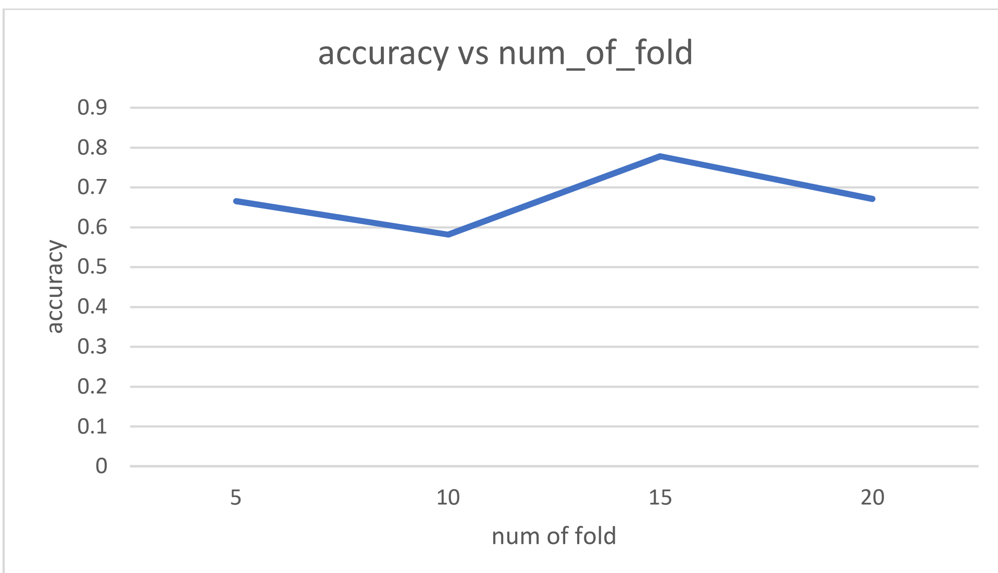
Part B

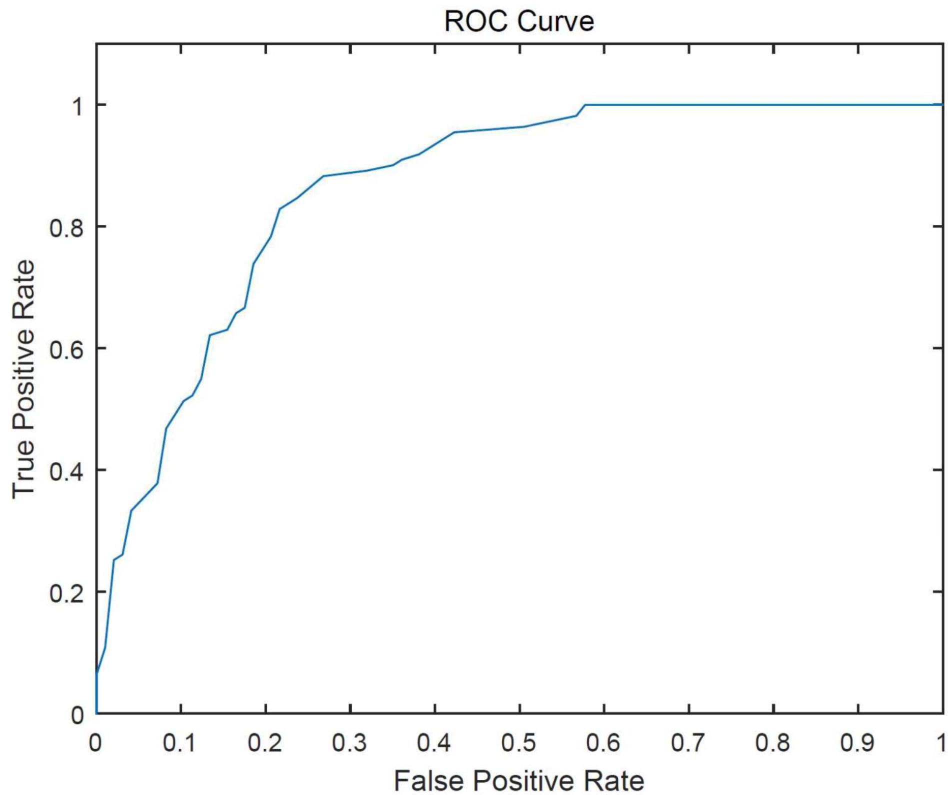
1.


accuracy vs num_of_epoch
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
accuracy
5
10
15
20
num of epoch

2.


accuracy vs num_of_fold
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
accuracy
5
10
15
20
num of fold

3.



Part C

1.Support : Binary vectors of length K

$x \in \{0, 1\}^K$

$Y \sim \text{Bernoulli}(ø)$

$X_k \sim \sim \text{Bernoulli}(\theta_{k,Y})\ \forall k \in \{1,\ldots\ldots,K\}$

Model: $p_{ø,\theta}(x, y) = (ø)^y(1 - ø)^{(1-y)}$

$$Ø = \frac{\sum_{i=1}^{N} \prod(y^{(i)} = 1)}{N}$$

$$\theta_{k,0} = \frac{\sum_{i=1}^{N} \prod(y^{(i)} = 0 \wedge x_k{}^{(i)} = 0\ )}{\sum_{i=1}^{N} \prod(y^{(i)} = 0)}$$

$$\theta_{k,1} = \frac{\sum_{i=1}^{N} \prod(y^{(i)} = 1 \wedge x_k{}^{(i)} = 1\ )}{\sum_{i=1}^{N} \prod(y^{(i)} = 1)}$$

$\forall k \in \{1, \ldots\ldots, K\}$

2.

According to the information showed in the problem. We can define

$P(Y = y_k|X) \propto \exp(w_{k0} + \sum_{i=1}^{d} w_{ki}X_i)$ for $k = 1, \ldots, K - 1$

Since all probabilities must sum to 1, we should have

$P\ (Y = y_k|X) = 1 - \sum_{k=1}^{K} P(Y = \ y_k|X)$

In binary classification, we define

$P(Y = y_k|X) = \frac{1}{1+ \sum_{k=1}^{K} \exp(w_{k0} + \sum_{i=1}^{d} w_{ki} X_i)}$

and for k = 1,……,K-1

$P(Y = y_k|X) = \frac{\exp(w_{k0} + \sum_{i=1}^{d} w_{ki} X_i)}{1+ \sum_{k=1}^{K} \exp(w_{k0} + \sum_{i=1}^{d} w_{ki} X_i)}$

Hence, it is clearly that p(y = i|x) is a softmax