ECE4880J: Computer Vision July 1, 2022

Homework 4: Machine Learning

Instructor: Siheng Chen **Yiwen Yang**

Instruction

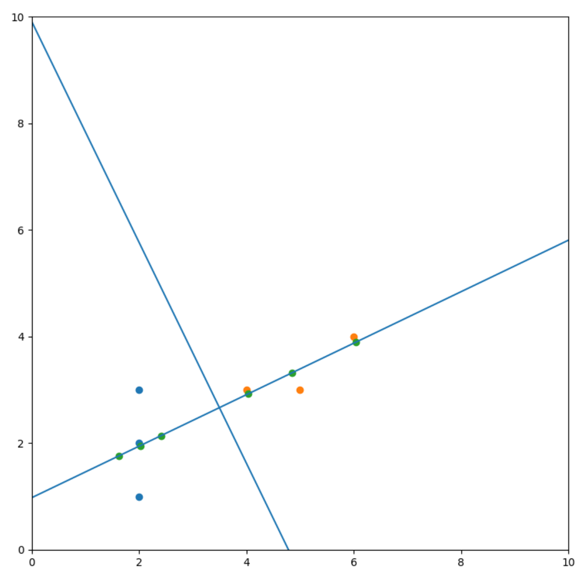
- This homework is due at 11:59:59 p.m. on ** July 11th, 2022.
- The write-up must be a soft copy `.pdf` file edited by LaTeX.
- The overall submission should be a `.zip` file named by xxx(student id)-xxx(name)-Assignment4.zip

Python Environment. We are using Python 3.7 for this course. We will use the following packages in this course: Numpy, Matplotlib.

Q1. Principal Component Analysis

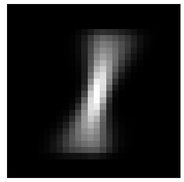
Part A

$MSE = 0.293$, $FR = 10.94$

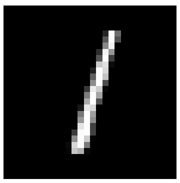


Part B

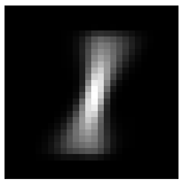
Using Gram Matrix trick takes 1.34s, while without using it takes 0.18s. It does not make sense to use Gram here because the data dimension is smaller than the sample number in this dataset.



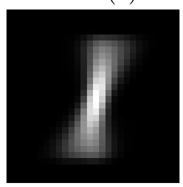
(a) Average one



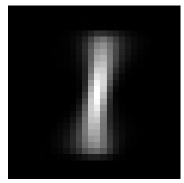
(b) Original



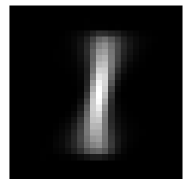
(c) n=1, MSE=2243691



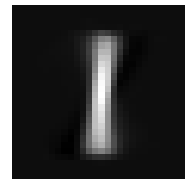
(d) n=2, MSE=1207623



(e) n=5, MSE=971508

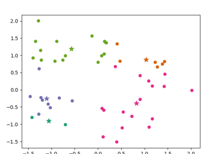


(f) n=10, MSE=856992

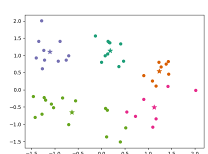


(g) n=20, MSE=855832

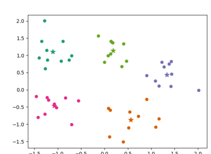
Q2. K-means



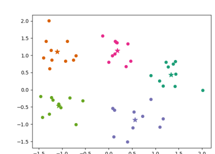
(a) c=5, i=1



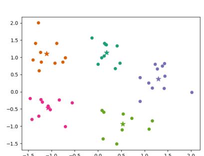
(b) c=5, i=2



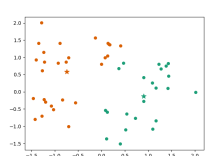
(c) c=5, i=4



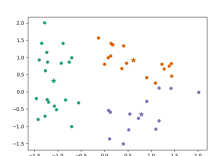
(d) c=5, i=8



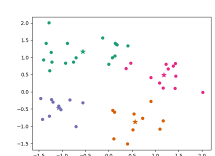
(e) c=5, i=16



(f) c=2, i=16



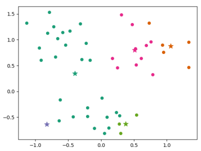
(g) c=3, i=16



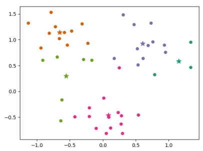
(h) c=4, i=16

Figure 2: Array 1

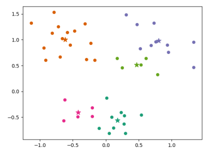
When cluster count K is fixed, the first iteration gives very random result, sometimes giving a centre that nobody belongs to; When iteration number grows, the centres tend to converge and almost not moving when $i = 16$. When iteration count i is fixed, it is observed that 2, 3 or 4 clusters does not fit array 1 very well, so array 1 should contains 5 clusters; However, 5 clusters does not converge in array 2, yet 3 clusters fit much better.



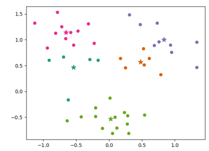
(a) c=5, i=1



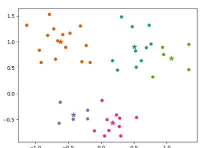
(b) c=5, i=2



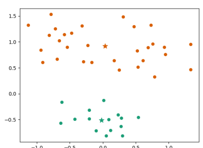
(c) c=5, i=4



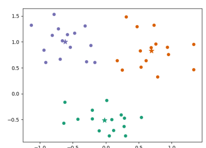
(d) c=5, i=8



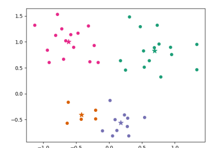
(e) c=5, i=16



(f) c=2, i=16



(g) c=3, i=16

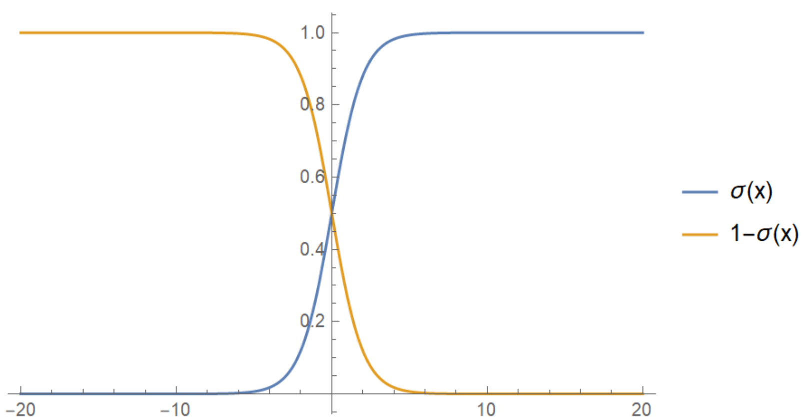


(h) c=4, i=16

Figure 3: Array 2

Q3. Logistic Regression

Part A

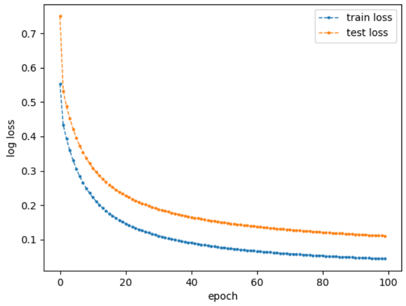


$1 - \sigma(x)$ is $\sigma(x)$ flipping with respect to the y-axis, i.e. $\sigma(-x) = 1 - \sigma(x)$.

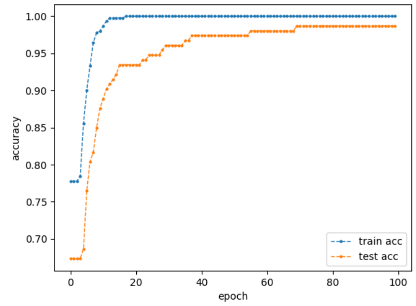
$$1 - \sigma(x) = 1 - \frac{1}{1+e^{-x}} = \frac{e^{-x}}{1+e^{-x}} = \frac{1}{1+e^{x}} = \sigma(-x)$$

Part B

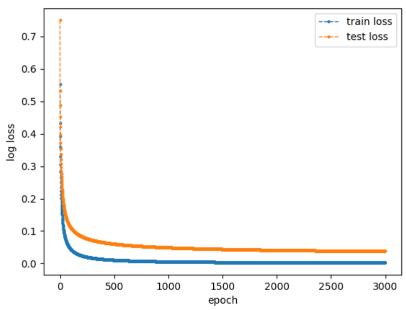
Test accuracy improves slightly at 500 and does not go any further, which implies that gradient converges around 500.



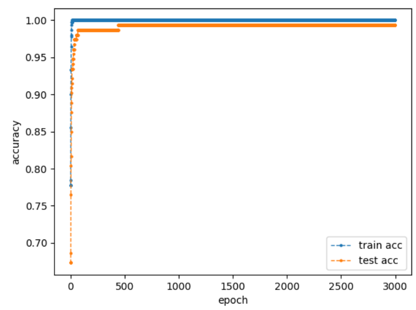
(a) 100-epoch loss



(b) 100-epoch accuracy



(c) 3000-epoch loss



(d) 3000-epoch accuracy

Q4. Linear Regression

1. Let

$$\frac{d}{d(\omega)} f(\omega) = 0$$

$$2(X^T\omega - y) \cdot X + 2\lambda \cdot \omega = 0$$

$$(X^T X + \lambda I)\omega = X^T y$$

$$\omega = (X^T X + \lambda I)^{-1} X^T y$$

2. $\lambda = 0$: $\omega = [4.44e - 16, 1]$

3. $\lambda = 1$: $\omega = [0.3636, 0.6591]$
   $\lambda = 1e - 3$: $\omega = [0.0016, 0.9986]$
   $\lambda = 1e - 5$: $\omega = [1.6e - 5, 0.9999]$

4.

$$J(\theta) = \sum_{i=1}^{N} \frac{1}{2}(y^{(i)} - \theta^T x^{(i)})^2$$

$$\nabla_\theta J(\theta) = \sum_{i=1}^{N} (\theta^T x^{(i)} - y^{(i)}) x^{(i)}$$

5. $i = 10$: $\omega = [0.4697, 0.5789]$
   $i = 100$: $\omega = [0.3637, 0.6888]$
   $i = 1000$: $\omega = [0.0231, 0.9802]$
   $i = 10000$: $\omega = [2.39e - 14, 1.0]$

6. Using L2 regularization and stopping early in gradient descent both produce an imperfect result but prevents overfitting.