Applied-Mathematics-for-CS-HW2

lizhaoyi SA22011035

December 8, 2022

1 Calculation Problems

1.1

solve for the following system of linear equations:

$$\begin{cases} x_1 - 2x_2 + 3x_3 - x_4 - x_5 = 2 \\ x_1 + x_2 - x_3 + x_4 - 2x_5 = 1 \\ 2x_1 - x_2 + x_3 - 0 - 2x_5 = 2 \\ 2x_1 + 2x_2 - 5x_3 + 2x_4 - x_5 = 5 \end{cases} \tag{1}$$

Solution 1:
The corresponding augmented matrix for this given system of linear equations is:

$$M_1 = \begin{bmatrix} 1 & -2 & 3 & -1 & -1 & 2 \\ 1 & 1 & -1 & 1 & -2 & 1 \\ 2 & -1 & 1 & 0 & -2 & 2 \\ 2 & 2 & -5 & 2 & -1 & 5 \end{bmatrix}$$

And we are going to utilize Gaussian-Elimination method to solve the above system of linear equation:
first round: we leverage the first row in the M_1 to eliminate the first element of the second row, the third row and the forth row.
we get the following transformed matrix M_2:

$$M_2 = \begin{bmatrix} 1 & -2 & 3 & -1 & -1 & 2 \\ 0 & 3 & -4 & 2 & -1 & -1 \\ 0 & 3 & -5 & 2 & 0 & -2 \\ 0 & 6 & -11 & 4 & 1 & 1 \end{bmatrix}$$

second round: Similarly, we leverage the second row in the M_2 to eliminate the second element of each row in the M_2, and reduce the second element in the second row to 1, then we get the following transformed matrix M_3:

$$M_3 = \begin{bmatrix} 1 & 0 & \frac{1}{3} & \frac{1}{3} & -\frac{5}{3} & \frac{4}{3} \\ 0 & 1 & -\frac{4}{3} & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & -1 & 0 & 1 & -1 \\ 0 & 0 & -3 & 0 & 3 & 3 \end{bmatrix}$$

third round: the same story, we leverage the third row in the M_3 to eliminate the third element of each row in the M_3, and reduce the third element in the third row to 1, then we get the following transformed matrix M_4:

$$M_4 = \begin{bmatrix} 1 & 0 & 0 & \frac{1}{3} & -\frac{4}{3} & 1 \\ 0 & 1 & 0 & \frac{2}{3} & -\frac{5}{3} & 1 \\ 0 & 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 6 \end{bmatrix}$$

conclusion: observe the matrix M_4 we can induce the rank of the augmented matrix $M_1 = rank(M_1) = rank(M_4) = 4$. Meanwhile, given that we are always conducting row-based elementary-transformation on

matrixs, so we can induce that the rank of the coefficient-matrix is $rank(M_1[:, :6]) = rank(M_4[:, :6]) = 3 < 4 = rank(M_1)$.
This implies that the above system of linear equations have no solution, for the rank of coefficient-matrix is strictly less than the rank of the corresponding augmented matrix.

1.2

for which values of parameters a and b, the following system of linear equations have solution? give that solution(s).

$$\begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 = 1 \\ 3x_1 + 2x_2 + x_3 + x_4 - 3x_5 = a \\ x_2 + 2x_3 + 2x_4 + 6x_5 = 3 \\ 5x_1 + 4x_2 + 3x_3 + 3x_4 - x_5 = b \end{cases} \tag{2}$$

Solution 2:
The corresponding augmented matrix for this given system of linear equations is:

$$M_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & 1 & -3 & a \\ 0 & 1 & 2 & 2 & 6 & 3 \\ 5 & 4 & 3 & 3 & -1 & b \end{bmatrix}$$

And we are going to utilize Gaussian-Elimination method to solve the above system of linear equation:
first round: we leverage the first row in the M_1 to eliminate the first element of the second row, the third row and the forth row.
we get the following transformed matrix M_2:

$$M_2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & -1 & -2 & -2 & -6 & a-3 \\ 0 & 1 & 2 & 2 & 6 & 3 \\ 0 & -1 & -2 & -2 & -6 & b-5 \end{bmatrix}$$

second round: Similarly, we leverage the second row in the M_2 to eliminate the second element of each row in the M_2, and reduce the second element in the second row to 1, then we get the following transformed matrix M_3:

$$M_2 = \begin{bmatrix} 1 & 0 & -1 & -1 & -5 & a-2 \\ 0 & 1 & 2 & 2 & 6 & 3-a \\ 0 & 0 & 0 & 0 & 0 & a \\ 0 & 0 & 0 & 0 & 0 & b-a-2 \end{bmatrix}$$

Principle: if we want the above system of linear equations do have at least one solution, we need to guarantee that the rank of the coefficient-matrix is equal to the rank of the corresponding augmented-matrix.
which means that:

$$\begin{cases} a = 0 \\ b - a - 2 = 0 \end{cases} \rightarrow \begin{cases} a = 0 \\ b = 2 \end{cases}$$

Meanwhile we can observe that when both of a = 0 and b = 2 exist:

$$rank(M_1) = rank(M_1[:, :6]) = 2 < 5, \text{ the number of variables} \tag{3}$$

So we know that the above system of linear equations have infinity solutions, and the general form of its solution is:

$$\begin{cases} x_1 = x_3 + x_4 + x_5 + 2 \\ x_2 = 3 - 2x_3 - 2x_4 - 6x_5 \\ x_3, x_4, x_5 \ can\ take\ any\ value \end{cases}$$

1.3

Supposed that there are 4 observed points which are (1,1), (2,1), (3,2) and (3,3), Utilize 2 methods of the least squared method to solve linear expression y = ax + b, making it fit these distributed points as good as possible.

Solution 3:

(1):normal method

According to the formula of least squared method, we know that:

$$a = \frac{n\Sigma_{i=1}^{n} x_i y_i - \Sigma_{i=1}^{n} x_i \Sigma_{i=1}^{n} y_i}{n\Sigma_{i=1}^{n} x_i^2 - (\Sigma_{i=1}^{n})^2}$$
$$b = \frac{1}{n}\Sigma_{i=1}^{n} y_i - \frac{a}{n}\Sigma_{i=1}^{n} x_i$$

where n is the total number of the observed data points.
according to the description of this problem, we know that:

$$(x_1, y_1) = (1,1),\ (x_2, y_2) = (2,1),\ (x_3, y_3) = (3,2),\ (x_4, y_4) = (3,3)$$

then with the above formula, we can calculate that:

$$\begin{cases} a = \frac{9}{11} \\ b = -\frac{1}{11} \end{cases}$$

with the 4 observed data points and least square method, we can calculate that the function of the sought line is:

$$y = \frac{9}{11}x - \frac{1}{11}$$

(2):matrix-based method

$$y = ax + b = (a, b) \cdot (x, 1)$$

let X denote the input features, Y denote the output labels.

$$X = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix};\ Y = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 3 \end{bmatrix}$$

So we know that the parameters are:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (X^T \cdot X)^{-1} X^T y = (\begin{bmatrix} 1 & 2 & 3 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 3 & 1 \end{bmatrix})^{-1} \begin{bmatrix} 1 & 2 & 3 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{9}{11} \\ -\frac{1}{11} \end{bmatrix}$$

So we can induce that the function of the sought line is:

$$y = \frac{9}{11}x - \frac{1}{11}$$

A	B	C	D
1	5	3	1
4	2	6	3
1	4	3	2
4	4	1	1
5	5	2	3

Figure 1: given data for problem4

1.4

Leverage PCA algorithm to reduce the dimension of given data to 2 and present your detailed calculation process:

Solution 4:

first let matrix X represent the originally given data's feature matrix.

$$X = \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ \dots & & & & \\ \dots & & & & \\ d_1 & d_2 & d_3 & d_4 & d_5 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 1 & 1 & 5 \\ 5 & 2 & 4 & 4 & 5 \\ 3 & 6 & 3 & 1 & 2 \\ 1 & 3 & 2 & 1 & 3 \end{bmatrix}$$

then we make the mean of each row to be **0**, we get that:

$$\overline{X} = \begin{bmatrix} -2 & 1 & -2 & 1 & 2 \\ 1 & -2 & 0 & 0 & 1 \\ 0 & 3 & 0 & -2 & -1 \\ -1 & 1 & 0 & -1 & 1 \end{bmatrix}$$

The covariance matrix of the above matrix is :

$$Cov(\overline{X}) = \begin{bmatrix} 2.8 & -0.4 & -0.2 & 0.8 \\ -0.4 & 1.2 & -1.4 & -0.4 \\ -0.2 & -1.4 & 2.8 & 0.8 \\ 0.8 & -0.4 & 0.8 & 0.8 \end{bmatrix}$$

We can calculate that the $\lambda_1 = 3.94$ and $\lambda_2 = 3.03$, where λ_1 and λ_2 are the largest 2 eigenvalue of Cov(X). And also, the corresponding eigenvector are:

$$p_1 = \begin{bmatrix} -0.26 & 0.91 & -0.31 & 0.09 \end{bmatrix}; \; p_2 = \begin{bmatrix} 0.48 & 0.04 & -0.51 & 0.71 \end{bmatrix}$$

$$P = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} -0.26 & 0.91 & -0.31 & 0.09 \\ 0.48 & 0.04 & -0.51 & 0.71 \end{bmatrix}$$

Consequently, we can get the reduced feature matrix Y:

$$Y = P \cdot X = \begin{bmatrix} 1.51 & -3.09 & 0.51 & 0.44 & 0.61 \\ -1.64 & -0.41 & -0.96 & 0.78 & 2.22 \end{bmatrix}$$

2 Calculation Problems

2.1

Dimension Reduction Method (PCA, AutoEncoder) and Classification:

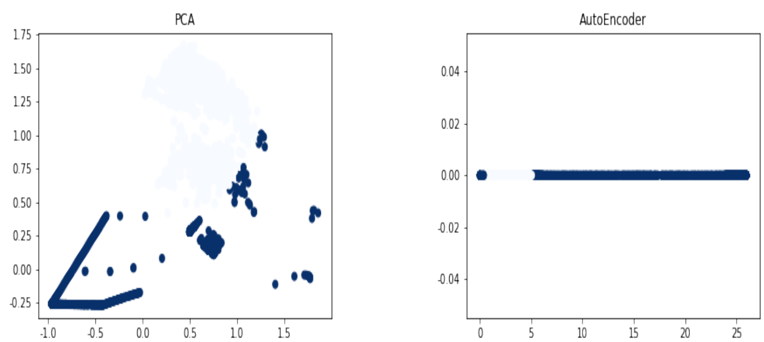
Report1:

(reduce_cls.ipynb)
(1):Performances variation with the different feature-dimensions after conducting dimension-reduction
We leverage Naive-Bayesian classifier as our classifier:

Target Dimension 40	30	20	10	5	2	
PCA	92.58%	96.40%	98.52%	98.76%	99.31%	99.50%
AutoEncoder	98.29%	97.20%	98.57%	98.91%	98.66%	99.23%

Observation:
PCA: the smaller the target dimension the higher the performance.
AutoEncoder: the Performance does not show obvious correlationship with the target dimension, while the performance can always be good(gearter than 97%).

(2):Plotings of 2-D distribution after conducting dimension-reduction (PCA, AutoEncoder) for the program runs for a long time to draw the whole picture, we **sample 50000 points** from the whole dataset to draw the illustration.

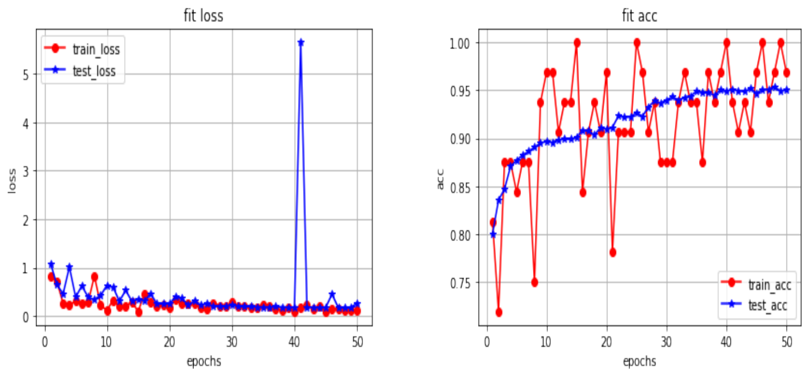


(a) (b)

Figure 2:

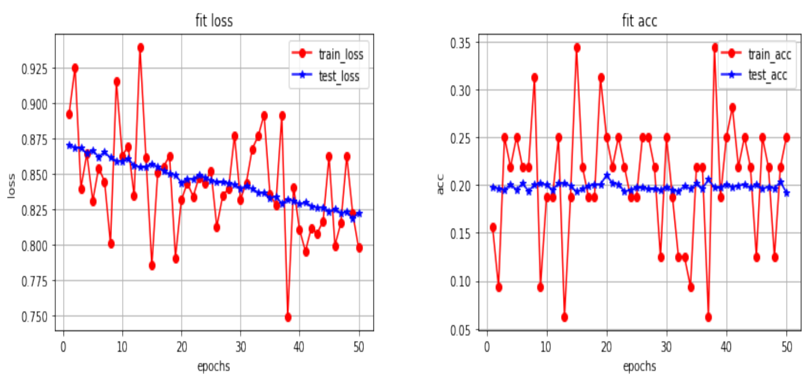
2.2

Skills to train a Multi-layer Neural Network
Report2:
(deep_train.ipynb)
(1)Model description:
We use a **5-layer** MLP as our classification model. From layer1 to layer 5, the numbers of the neurons in each layer are **41, 36, 24, 12, 6, 1**, correspondingly. We set 2 kind of activation fuctions: **relu** or **tanh**. Besides, we try 2 kind of different initialization methods of model's parameters, which are **Xaiver's initialization** and **normal distribution** with mean = 0 and std = 0.001. Finally, we leverage off-the-shelf **SGD** and **Adam** optimizer in PyTorch1.1 to optimize our model.
(2)Training-Testing Losses and Acces variation with epochs:
My Overall Conclusion and Observation:
In our case and model architecture setting, I found that with **relu-activation**, **Xaiver-initialization**, **Adam-optimizer**, our model can be trained more steadily and meanwhile achieve better performance with a faster convergence speed.
Here are some poltings to illustrate our observations:



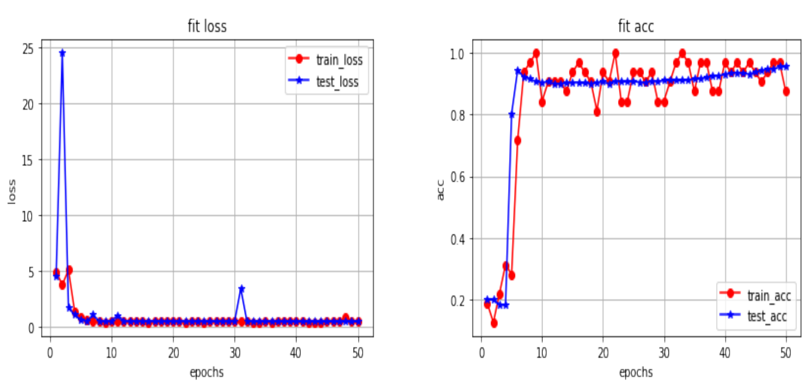
(a) (b)

Figure 3: SGD-relu



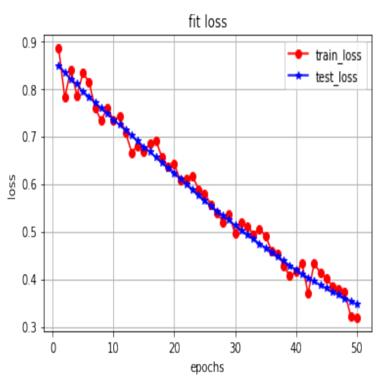
(a) (b)

Figure 4: SGD-tanh

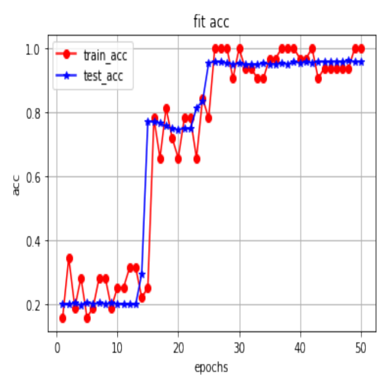


(a) (b)

Figure 5: Adam-relu



(a)



(b)

Figure 6: Adam-tanh