CSCE 636: Deep Learning

Fall 2021 – HW 3

1. Given,

A is 3x3 square matrix, whose eighen value decomposition is given by:

$$A = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix} \qquad (i)$$

$$Or, A = u\Lambda u^T$$

We know that for a given square matrix (S), the eigen value decomposition is given by:

$$S = Q\Lambda Q^{-1} \qquad (ii)$$

Therefore, from (i) and (ii), we get that $u^T = u^{-1}$, i.e. u is an orthogonal matrix.

And, we also know that, for any given matrix (R):

$$R.R^T = U \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} U \qquad (iii)$$

And, $\qquad (iv)$

$$R^T.R = V\,\Sigma^2 V^T$$

Using (iii), we have:

$$A.A^T = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix} . \left(\begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix} \right)^T$$

Or,

$$A.A^T = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}^T \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}^T \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix}^T$$

Or, $$A.A^T = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} 1 \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}^T \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix}^T \quad Since, u^T = u^{-1}$$

Or, $$A.A^T = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}$$

$$\text{Or,} \qquad A.A^T = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}$$

$$\text{Or,} \qquad A.A^T = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{31} & u_{32} & u_{33} \\ u_{21} & u_{22} & u_{23} \end{bmatrix} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{31} & u_{32} & u_{33} \\ u_{21} & u_{22} & u_{23} \end{bmatrix}$$

$$\text{Therefore,} \qquad U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{31} & u_{32} & u_{33} \\ u_{21} & u_{22} & u_{23} \end{bmatrix}$$

And, Using (iv), we have:

$A^T.A$

$$= \left(\begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix} \right)^T . \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}$$

$Or, A^T.A$

$$= \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}^T \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}^T \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix}^T . \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}$$

$$Or, A^T.A = \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}^T \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}^T 1 \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix} \qquad Since, u^T = u^{-1}$$

$$Or, A^T.A = \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}^T \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}$$

$$Or, A^T.A = \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}^T \begin{bmatrix} 9 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}$$

$$Or, A^T.A = \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}^T$$

$$Or, A^T.A = \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}^{-1} \qquad Since, u^T = u^{-1}$$

$$Or, A^T.A = \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{13} & u_{23} & u_{33} \\ u_{12} & u_{22} & u_{23} \end{bmatrix} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{13} & u_{23} & u_{33} \\ u_{12} & u_{22} & u_{23} \end{bmatrix}^{-1}$$

$$\text{Therefore,} \qquad V = \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{13} & u_{23} & u_{33} \\ u_{12} & u_{22} & u_{23} \end{bmatrix}$$

And,

$$\Sigma = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Hence, as per definition of singular value decomposition, we have:

$$A = U\Sigma V^T$$

$$Or, \ A = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{31} & u_{32} & u_{33} \\ u_{21} & u_{22} & u_{23} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{13} & u_{23} & u_{33} \\ u_{12} & u_{22} & u_{23} \end{bmatrix}^T$$

$$Or, \ A = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{31} & u_{32} & u_{33} \\ u_{21} & u_{22} & u_{23} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{13} & u_{12} \\ u_{21} & u_{23} & u_{23} \\ u_{31} & u_{33} & u_{23} \end{bmatrix}$$

2. Given,

H is a symmetric (nxn) matrix with:

- eigenvalues: $\lambda_1 \geq \lambda_{2,} \geq \lambda_3 \ldots. \geq \lambda_n$
- eigenvectors: $U = [u_1, \, u_2, \ldots\ldots, u_n \,]$

Therefore, the eigen value decomposition of H is given by

$$H = U\Sigma U^T$$

$$\text{where, } \Sigma = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & . & 0 & 0 \\ 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & \lambda_n \end{bmatrix}$$

Since, H is symmetric, U is orthogonal.

$$\text{Now, } A^T H A = A^T(U\Sigma U^T)A = (U^T A)^T \Sigma (U^T A)$$

$$\text{Let, } (U^T A) = B, such\ that\ B^T B = I$$

$$\text{Therefore, } tr(A^T H A) = tr\ (B^T \Sigma B) = tr(\Sigma B B^T) = \textstyle\sum_{i=1}^{n} \lambda_i \left||b_i|\right|^2 \qquad (i)$$

Since, non-diagonal elements of Σ are 0

Now, we know that $B^T B = I$

Let us represent B as a matrix consisting of n rows: $b_1, b_2, \ldots . b_n$.

Therefore, B^T would then be a matrix consisting of n columns: $b_1, b_2, \ldots . b_n$.

Hence, $B^T B = \begin{bmatrix} {b_1}^2 & b_2 b_1 & . & . & b_n b_1 \\ b_1 b_2 & {b_2}^2 & . & . & b_n b_2 \\ . & . & . & . & . \\ . & . & . & . & . \\ b_1 b_n & b_2 b_n & . & . & {b_n}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & . & . & 0 \\ 0 & 1 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & 1 \end{bmatrix}$

Therefore,

$$ {b_1}^2 = {b_2}^2 = \cdots = {b_n}^2 = 1 \qquad (ii) $$

Hence from (i) and (ii), we have:

$$ tr(A^T H A) = \sum_{i=1}^{k} \lambda_i $$

[Hence, Proved]

3. You may find the completed code in the submitted files.
   Please find the results of the hyperparameter training (highlighted in yellow) and testing accuracy below:

- Kernel_LR

Sl. No.	Hyperparameter Combo	Average Validation Acc
1	• learning_rate = 0.01 • max_epoch = 10 • batch_size = 64 • sigma = 1	0.80848
2	• learning_rate = 0.01 • max_epoch = 10 • batch_size = 64 • sigma = 2	0.97156
3	• learning_rate = 0.001 • max_epoch = 10 • batch_size = 64 • sigma = 2	0.98545

Testing accuracy - **0.9502**

- RBF

Sl. No.	Hyperparameter Combo	Average Validation Acc

1	• hidden_dim = 200 • learning_rate = 0.01 • max_epoch = 10 • batch_size = 64 • sigma = 2	0.9344
2	• hidden_dim = 300 • learning_rate = 0.01 • max_epoch = 10 • batch_size = 64 • sigma = 2	0.9527
3	• hidden_dim = 300 • learning_rate = 0.001 • max_epoch = 10 • batch_size = 64 • sigma = 2	0.9653

Testing accuracy - **0.9329**

- FFN

Sl. No.	Hyperparameter Combo	Average Validation Acc
1	• hidden_dim = 64 • learning_rate = 0.01 • max_epoch = 10 • batch_size = 64	0.9979
2	• hidden_dim = 128 • learning_rate = 0.01 • max_epoch = 10 • batch_size = 64	0.9997
3	• hidden_dim = 300 • learning_rate = 0.001 • max_epoch = 10 • batch_size = 64	0.9990

Testing accuracy - **0.98268**

4. You may find the completed code in the submitted files.
   Please find the results below:

- Reconstruction Error

d	PCA Reconstruction Error	Autoencoder Reconstruction Error (Sigmoid, Tanh, ReLU)
32	269.4172	178.4242, 494.5225, 410.2892
64	251.5158	142.8160, 515.9313, 417.6237
128	240.8621	111.6051, 633.5126, 430.8327

The Autoencoder seems to better reconstruct the inputs. This may be because autoencoder involved some kind of learning (where the error/loss to reconstruct the inputs is targeted to decrease at every iteration). However, in PCA the reconstruction simply takes place based on reduced rank approximation of the matrix. And it is observed that the higher the hidden dimension, the smaller is the reconstruction error in both cases.

- Comparing W and G

d	Frobeniu Norm between W and G (Sigmoid, Tanh, ReLU)
32	75.9503, 88.3314, 90.4067
64	95.8757, 126.1780, 127.5083
128	143.4463, 158.6712, 179.8638

Here, it is observed that as the hidden dimension increases, the differences between the PCA weights and Autoencoder weights also increases.

- Autoencoder Reconstruction Error without weights sharing

d	Weights sharing Reconstruction Error (Sigmoid, Tanh, ReLU)	Weights not sharing Reconstruction Error (Sigmoid, Tanh, ReLU)
32	178.4242, 494.5225, 410.2892	170.9591, 405.6484, 386.2084
64	142.8160, 515.9313, 417.6237	148.0365, 413.5284, 404.9223
128	111.6051, 633.5126, 430.8327	102.0406, 627.3070, 404.8511

The performance seems marginally better when there is no sharing of weights.

- Autoencoder Reconstruction Error without with multiple layers

Model Parameters	Reconstruction Error
2 hidden layers in encoder and decoder. ReLU Activation. (Encoder size: 256-192-128-64. Decoder size: 64-128-192-256)	128.0155
1 hidden layers in encoder and decoder. ReLU Activation (Encoder size: 256-128-64. Decoder size: 64-128-256)	124.2723
2 hidden layers in encoder and decoder. Sigmoid Activation. (Encoder size: 256-192-128-64. Decoder size: 64-128-192-256)	152.2071
1 hidden layers in encoder and decoder. Sigmoid Activation (Encoder size: 256-128-64. Decoder size: 64-128-256)	112.9187
2 hidden layers in encoder and decoder. Tanh Activation. (Encoder size: 256-192-128-64. Decoder size: 64-128-192-256)	88.9893
1 hidden layers in encoder and decoder. Tanh Activation (Encoder size: 256-128-64. Decoder size: 64-128-256)	94.6996

For the multiple layers approach, tanh seemed to be the best activation function. And also, intuitively, the greater number of layers lead to smaller reconstruction error.