AutoEncoders (AE) & Principal Component Analysis (PCA)

African Master in Machine Intelligence, AMMI

Foundations of ML & DL

Supervised by : Prof. MOUSTAPHA Cisse

April 21, 2023

AIMS African Institute for Mathematical Sciences SENEGAL

Group 5 & 7's members

- Phanie
- El mamoune
- Regis
- Dieu-Donne
- John
- Mame Diara Diouf
- Khady Gaye

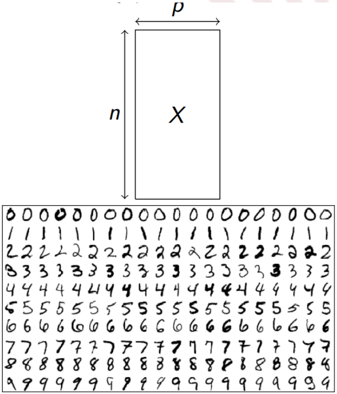
Overview

1. Motivation
2. Equivalence between PCA and LAEs
   - Equivalence between PCA and LAEs
   - Experimental Case: PCA vs LAEs
3. Denoising AutoEncoder
   - Denoising AutoEncoder
4. Contractive AutoEncoder
   - Contractive AutoEncoder
5. Conclusion
   - Conclusion
6. Reference

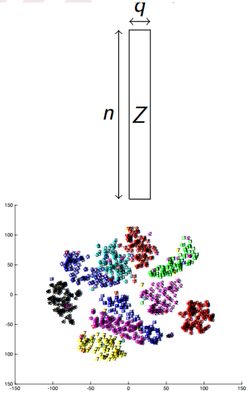
Motivation

About PCA

let X be the dataset with n individus and $p >> 2$ variable.



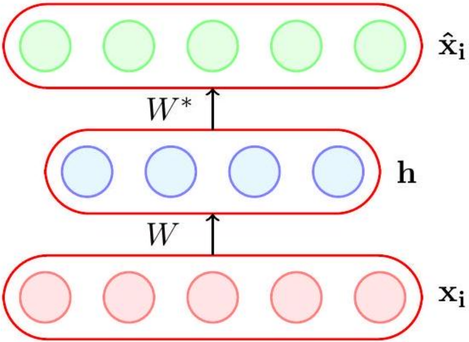
$p = 784$



$q = 2$

Motivation

About AutoEncoders



$$\mathbf{h} = g\left(W\mathbf{x_i} + \mathbf{b}\right)$$

$$\hat{\mathbf{x}}_\mathbf{i} = f\left(W^*\mathbf{h} + \mathbf{c}\right)$$

Overview

1. Motivation
2. Equivalence between PCA and LAEs
   - Equivalence between PCA and LAEs
   - Experimental Case: PCA vs LAEs
3. Denoising AutoEncoder
   - Denoising AutoEncoder
4. Contractive AutoEncoder
   - Contractive AutoEncoder
5. Conclusion
   - Conclusion
6. Reference

Equivalence between PCA and LAEs

We will now see that the encoder part of an autoencoders is equivalent to PCA if we:

- normalize the input to

$$\hat{x}_{ij} = \frac{1}{\sqrt{m}} \left(x_{ij} - \frac{1}{m} \sum_{k=1}^{m} x_{kj} \right) \tag{2.1}$$

- use a linear encoder;
- use a linear decoder ;
- use a MSE, loss function.

Equivalence between PCA and LAEs

First let us consider the implication of normalizing the inputs to

$$\hat{x}_{ij} = \frac{1}{\sqrt{m}}\left(x_{ij} - \frac{1}{m}\sum_{k=1}^{m} x_{kj}\right) \tag{2.2}$$

- The operation in the bracket ensures that the data now has 0 mean along each dimension j (we are subtracting the mean)
- Let X' be this zero mean data matrix then what the above normalization gives us is $X = \frac{1}{\sqrt{m}}X'$
- Now $(X)^T X = \frac{1}{m}(X')^T X'$ is the covariance matrix (recall that covariance matrix plays an important role in PCA)

Equivalence between PCA and LAEs

First we will show that if we use linear decoder and a squared error loss function then, the optimal solution to the following objective function

$$\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\hat{x}_{ij})^2 \tag{2.3}$$

is obtained when we use a linear encoder.
This is equivalent to

$$\min_{\theta}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\hat{x}_{ij})^2 \tag{2.4}$$

$$\min_{H;W^*}(\|X-HW^*\|_F)^2 \quad \|A\|_F=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}^2} \tag{2.5}$$

Equivalence between PCA and LAEs

Reminder: Singular Value Decomposition (SVD)

Theorem

For any matrix $X \in \mathbb{R}^{n\times k}$, there exist two orthogonal matrices $U \in \mathbb{R}^{n\times n}$, $V \in \mathbb{R}^{k\times k}$ and a nonnegative, "diagonal" matrix $\Sigma \in \mathbb{R}^{n\times k}$ (of the same size as X) such that

$$X_{n\times k} = U_{n\times n}\mathbf{\Sigma}_{n\times k}V^T_{k\times k} \tag{2.6}$$

From the SVD of X we obtain that

$$X^TX = V\mathbf{\Sigma}^TU^T \cdot U\mathbf{\Sigma}V^T = V\left(\mathbf{\Sigma}^T\mathbf{\Sigma}\right)V^T \tag{2.7}$$

This shows that V is the eigenvectors matrix of X^TX.

Equivalence between PCA and LAEs

- So, from SVD we know that the optimal solution to the optimization problem (2.5) is given by

$$HW^* = U\Sigma V^T \tag{2.8}$$

- By matching variables one possible solution is

$$\begin{aligned} H &= U\Sigma \\ W^* &= V^T \end{aligned} \tag{2.9}$$

- We will now show that H is a linear encoding and find an expression for the encoder weights W

Equivalence between PCA and LAEs

$$
\begin{aligned}
H &= U\Sigma \\
&= \left(XX^T\right)\left(XX^T\right)^{-1} U\Sigma \qquad ;(\text{Since } \left(XX^T\right)\left(XX^T\right)^{-1} = I\Big) \\
&= \left(XV\Sigma^T U^T\right)\left(U\Sigma V^T V\Sigma^T U^T\right)^{-1} U\Sigma \quad ;\Big(u\,\text{sing}\, X = U\Sigma V^T \ ) \\
H &= XV
\end{aligned}
$$

Thus H is a linear transformation of X and $W = V$

- We have encoder $W = V$
- From SVD, we know that V is the matrix of eigen vectors of $X^T X$
- Thus, the encoder matrix for linear autoencoder (W) and the projection matrix (P) for PCA could indeed be the same.

Hence proved

Overview

1. Motivation
2. Equivalence between PCA and LAEs
   - Equivalence between PCA and LAEs
   - Experimental Case: PCA vs LAEs
3. Denoising AutoEncoder
   - Denoising AutoEncoder
4. Contractive AutoEncoder
   - Contractive AutoEncoder
5. Conclusion
   - Conclusion
6. Reference

AIMS African Institute for Mathematical Sciences SENEGAL

Equivalence between PCA and LAEs

Experimental Case: PCA vs LAEs

- In this experiment we have considered IRIS data. IRIS data includes 150 × 4 data. It contains sepal width, sepal length, petal width, petal length as columns;

- We use this data to demonstrate relation between PCA and linear auto encoder. We standardise the IRIS data and then run PCA on it, later we apply auto encoder with one hidden layer with linear encoder, decoder with mean squared error:

$$L(X, \hat{X}) = \frac{1}{N} ||X - \hat{X}||_2^2$$

Equivalence between PCA and LAEs

Experimental Case: PCA vs LAEs

- In PCA we have:

$$\hat{X} = XW$$

Which W is the matrix formed by the principal components.

- in Linear Auto encoder we have:

$$Z = f(W_1X) \qquad \hat{X} = g(W_2Z)$$

were f and g are linear functions we have:

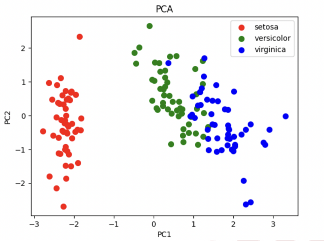
$$Z = W_1^{'}X \qquad \hat{X} = W_2^{'}Z$$

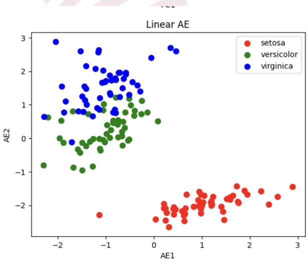
$$\hat{X} = W_2^{'}W_1^{'}X$$

Equivalence between PCA and LAEs

Experimental Case: PCA vs LAEs

Below is the plot for PCA vs Linear auto encoder on IRIS dataset.





Equivalence between PCA and LAEs

Experimental Case: PCA vs LAEs

- PCA and Linear Auto encoder fails terribly if the data is non linear, but auto-encoder works well there too to their ability to do non-linear transformations.
- Auto encoder is prone to overfitting due to high number of parameters (though regularization and careful design can avoid this), we will see in next section.

Overview

1. Motivation
2. Equivalence between PCA and LAEs
    - Equivalence between PCA and LAEs
    - Experimental Case: PCA vs LAEs
3. Denoising AutoEncoder
    - Denoising AutoEncoder
4. Contractive AutoEncoder
    - Contractive AutoEncoder
5. Conclusion
    - Conclusion
6. Reference

Denoising AutoEncoder

- DAEs are variants of AEs that are trained to reconstruct a clean and repared input from corrupted versions.
- It is done by corrupting the initial input vector x into a transformed $\tilde{x}$ by using a stochastic mapping $\tilde{x} \sim q_D(\tilde{x}|x)$, where q_D is a stochastically corrupted process.
- During the training, each example x is corrupted generating $\tilde{x}$ according to $q_D(\tilde{x}|x)$ in order to fit the model.
- Considering f and g respectively as encoder and decoder, the loss function is as follows:

$$L_D(x, \hat{x}) = \frac{1}{|D|} \sum_{x \in D} ||x - g(f(\tilde{x}))||^2, \textbf{ where } h = f(\tilde{x}) \textbf{ and } \hat{x} = g(h) \tag{5.1}$$

Architecture of a denoising autoencoder

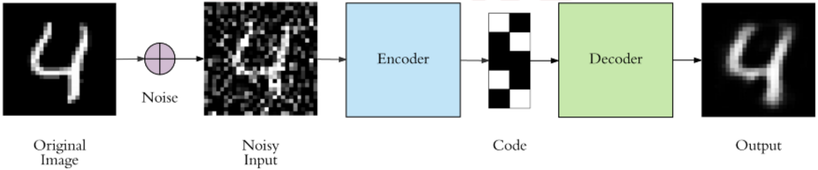


Figure: General structure of a denoising autoencoder

Overview

1. Motivation
2. Equivalence between PCA and LAEs
   - Equivalence between PCA and LAEs
   - Experimental Case: PCA vs LAEs
3. Denoising AutoEncoder
   - Denoising AutoEncoder
4. Contractive AutoEncoder
   - Contractive AutoEncoder
5. Conclusion
   - Conclusion
6. Reference

Contractive AutoEncoder

- A contractive autoencoder tries to prevent an overcomplete autoencoder from learning the identity function.
- It does it by adding an explicit regularization term to the loss function.
- The is goal is to extract features that only reflect variations observed in the training set. The model must be invariant to the other variations.

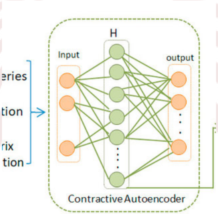


Figure: General structure of an overcomplete autoencoder

Loss function of a contractive autoencoder

- The loss function is defined as follows:

$$\begin{aligned} L_D(x) &= \sum_{x \in D} \mathcal{L}(x, g(f(x))) + \lambda ||J_f(x)||_F^2, \\ ||J_f(x)||_F^2 &= \sum_{i,j} (\frac{\partial h_j(x)}{\partial x_i})^2 \end{aligned} \tag{4.1}$$

- where $\mathcal{L}$ is the reconstruction error either MSE or CE loss for autoencoders
- $J_f(x) = \frac{\partial h}{\partial x}(x)$ is the regularization term which is the Jacobian of the hidden representation of the input x

AIMS African Institute for Mathematical Sciences SENEGAL

69

Overview

1. Motivation
2. Equivalence between PCA and LAEs
    - Equivalence between PCA and LAEs
    - Experimental Case: PCA vs LAEs
3. Denoising AutoEncoder
    - Denoising AutoEncoder
4. Contractive AutoEncoder
    - Contractive AutoEncoder
5. Conclusion
    - Conclusion
6. Reference

Conclusion

- Equivalence between PCA and LAEs is due to the fact that the optimization problem for the linear autoencoder is similar to the eigenvalue decomposition problem that is used to compute the principal components of the data.
- The idea behind denoising autoencoder is just to increase the robustness of the encoder to the small changes in the training data which is quite similar to the motivation of Contractive Autoencoder. However, there is some difference:

Conclusion

1. CAEs encourage robustness of representation encoder, whereas DAEs encourage robustness of reconstruction, which only partially increases the robustness of representation.

2. DAE increases its robustness by stochastically training the model for the reconstruction, whereas CAE increases the robustness of the first derivative of Jacobian matrix.

Reference

[1] Rifai, S., Vincent, P., Muller, X., Glorot, X., Bengio, Y. (2011, June). *Contractive auto-encoders: Explicit invariance during feature extraction. In Proceedings of the 28th international conference on international conference on machine learning (pp. 833-840).* , Dept. IRO, Université de Montréal. Montréal(QC), H3C 3J7, Canada

[2] Umberto Michelucci (2022) *An Introduction to Autoencoders* , TOELT.AI.

[3] Alireza Makhzani (2018) *Unsupervised Representation Learning with Autoencoders*, Ph.D. thesis, University of Toronto.

[4] William L. Hamilton (1994) *Fundamentals of Machine Learning, Lecture 25 — Autoencoders and self-supervision.*

AIMS African Institute for Mathematical Sciences SENEGAL

JËJËRËF