Reducing the Dimensionality of Data with Neural Networks

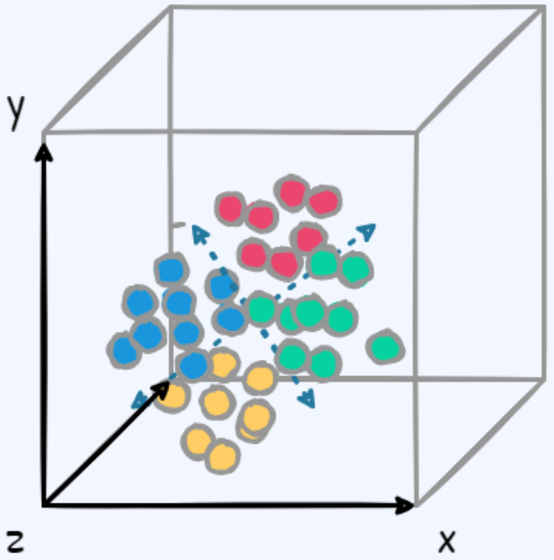
Donghwan Yang

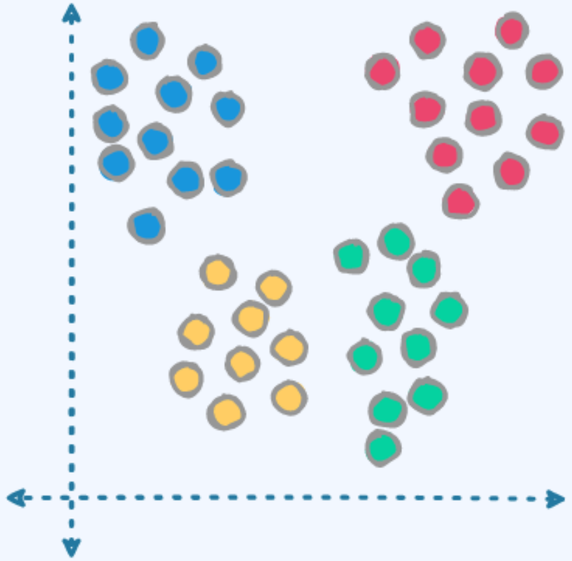
February 6, 2025

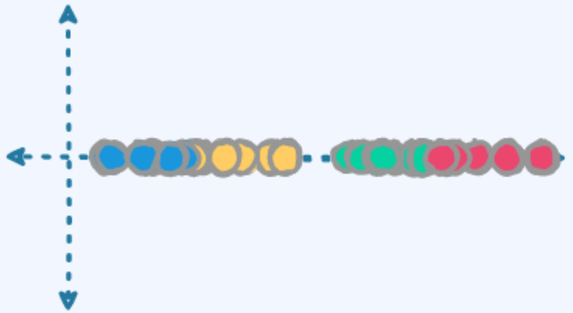
Contents

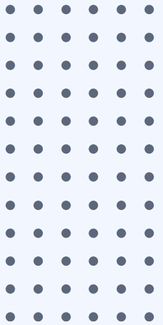
What is Dimensionality Reduction?

Dimensionality Reduction

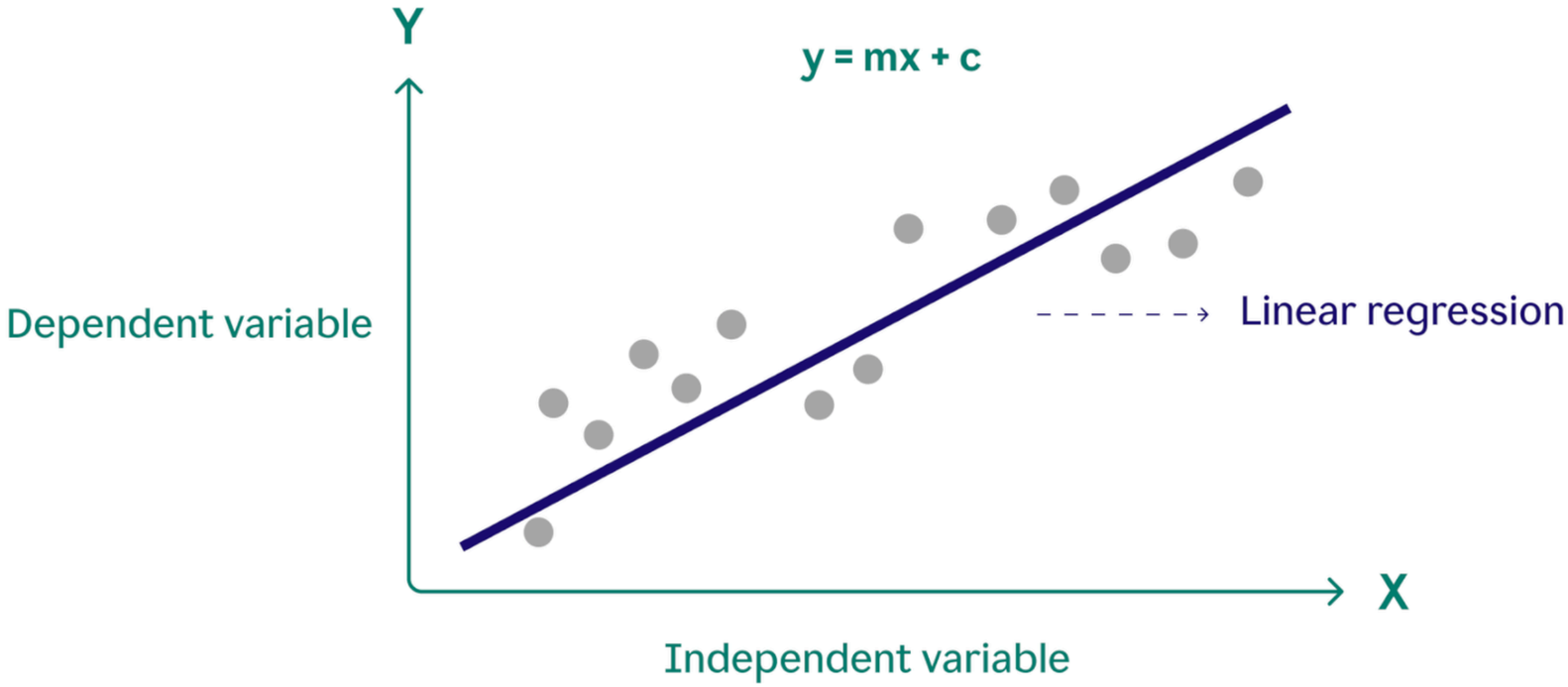








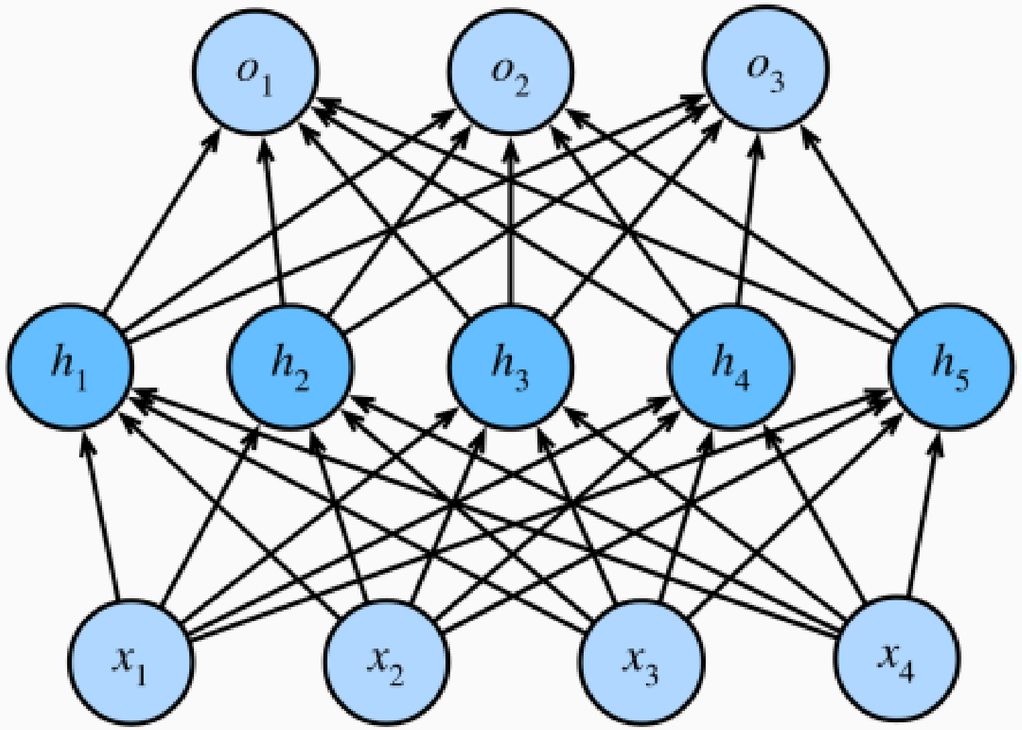
Artificial Neural Network


Y
y = mx + c
Dependent variable
Linear regression
X
Independent variable

Output layer

Hidden layer

Input layer

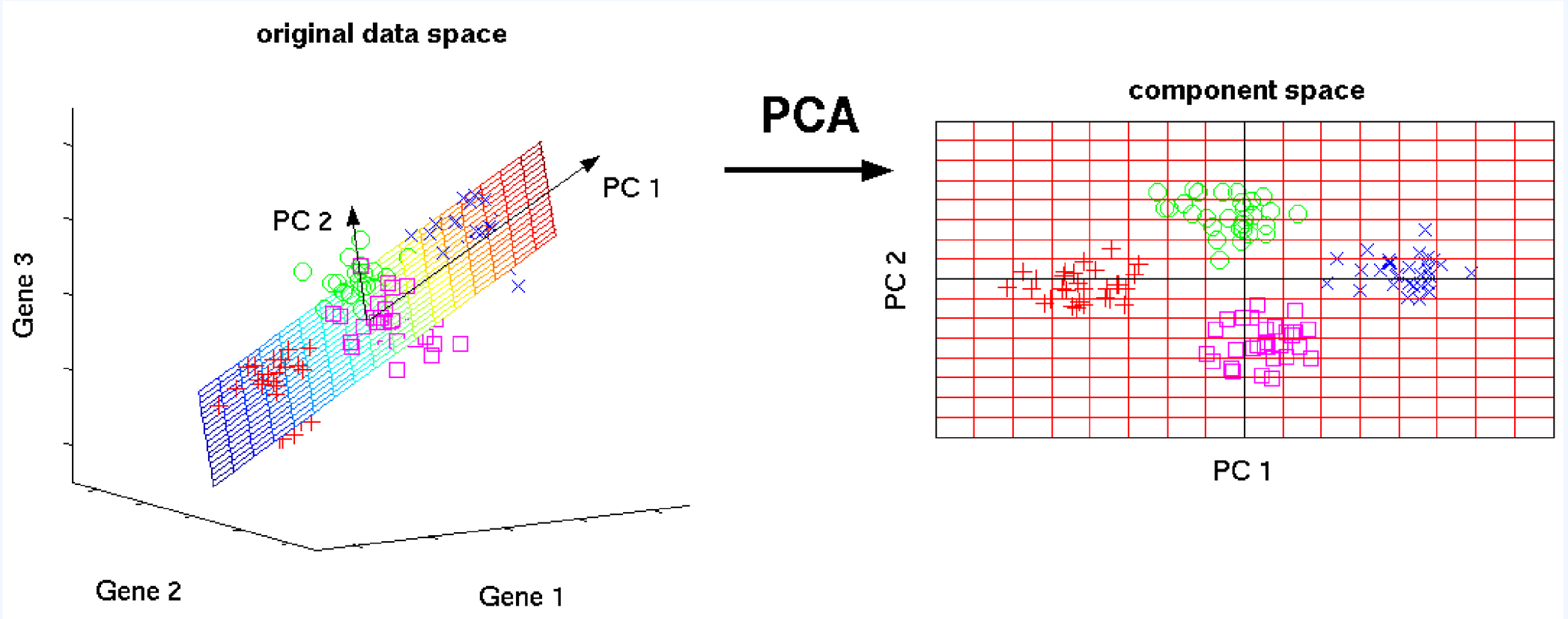


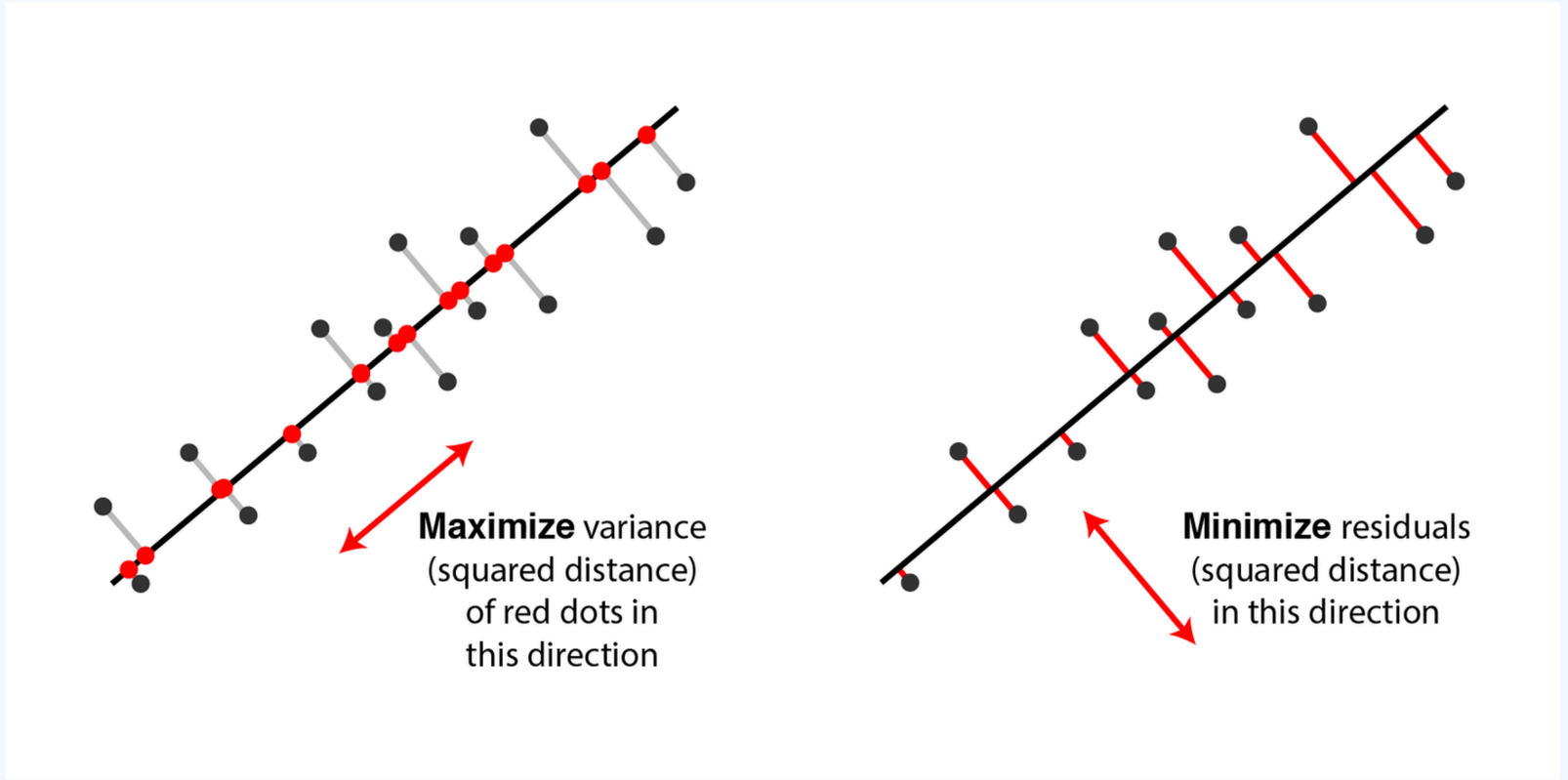
$\mathbf{W}^{(2)} \in \mathbb{R}^{h \times k}, \mathrm{b}^{(2)} \in \mathbb{R}^{1 \times k}$

$\mathbf{W}^{(1)} \in \mathbb{R}^{p \times h}, \mathrm{b}^{(1)} \in \mathbb{R}^{1 \times h}$

$\mathbf{X} \in \mathbb{R}^{n \times p}$

Principal Component Analysis


original data space
PC 1
PC 2
Gene 3
Gene 2
Gene 1
PCA
component space
PC 2
PC 1

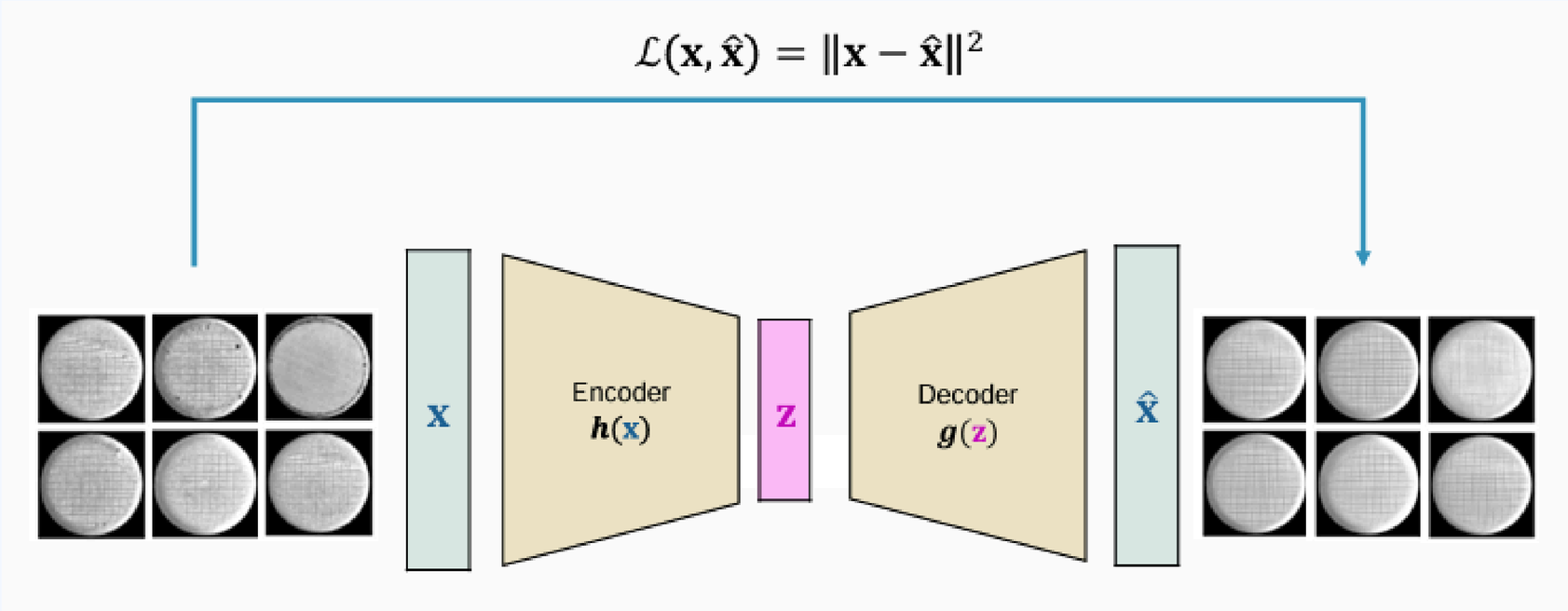

Maximize variance
(squared distance)
of red dots in
this direction
Minimize residuals
(squared distance)
in this direction

AutoEncoder

AutoEncoder

Cost function of AutoEncoder

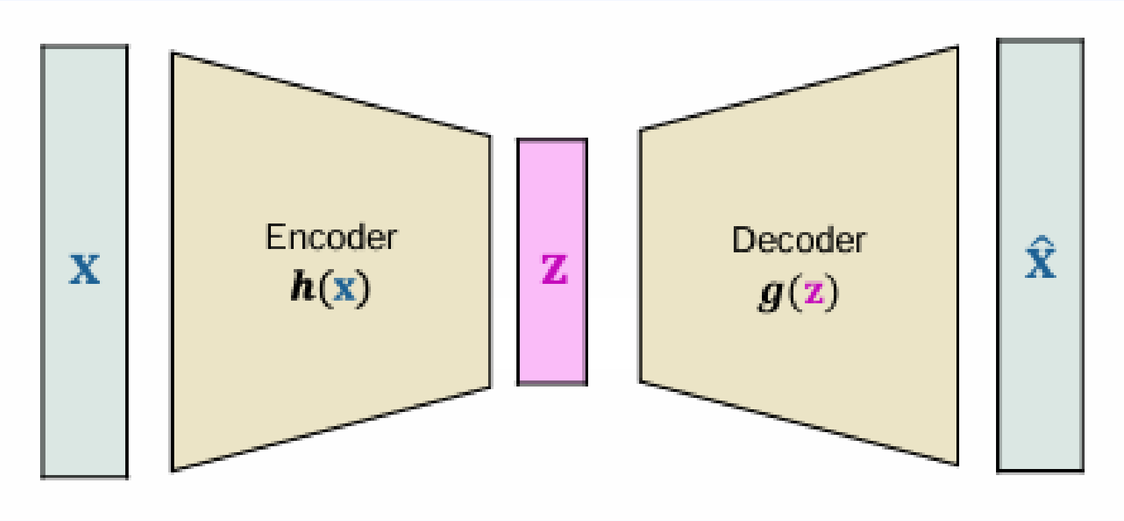
- Features should be extracted as useful information
- Mean squared error between the original and the reconstructed image (or signal)



AutoEncoder

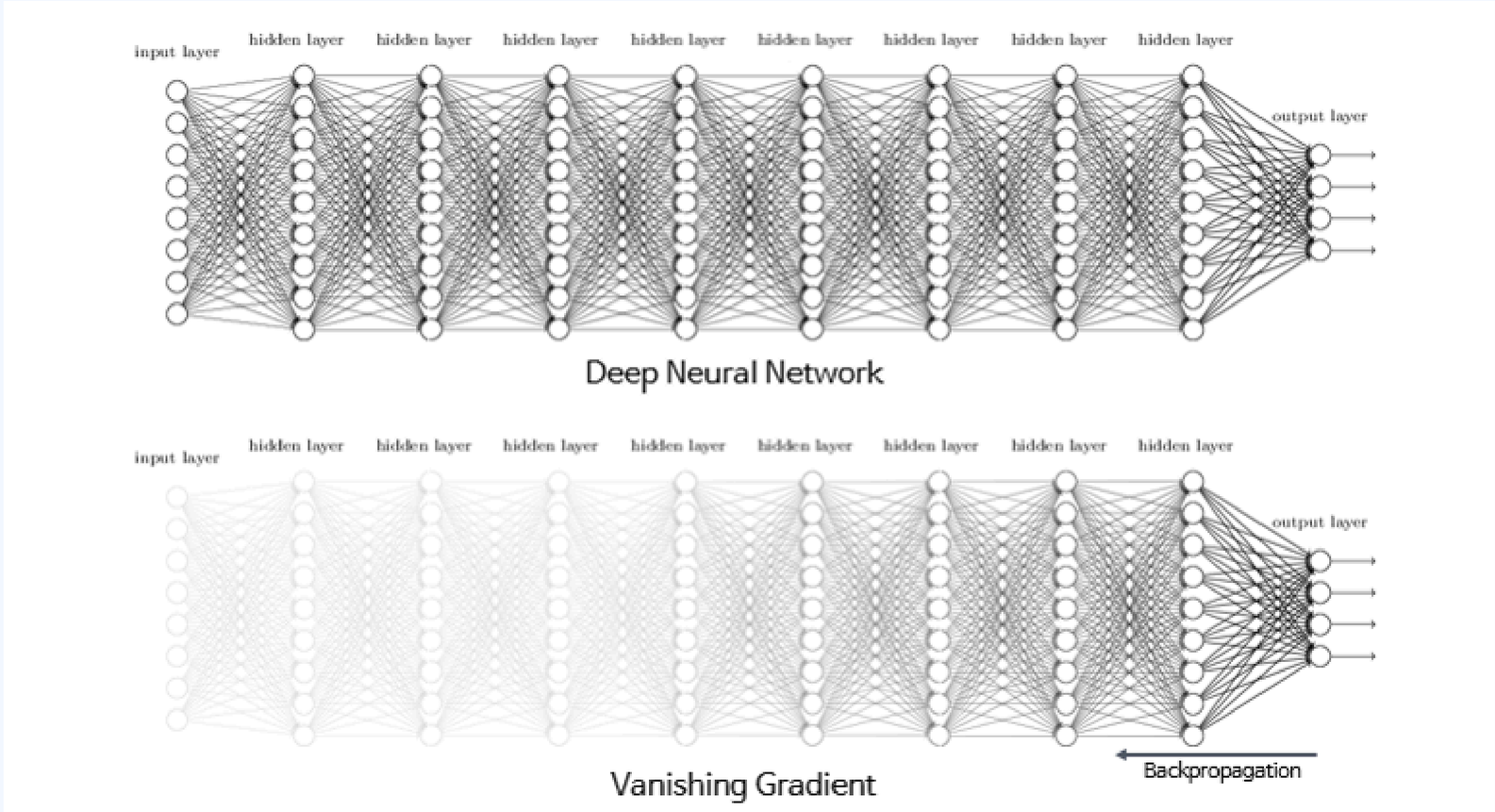
How to Use Autoencoders

- After training, the decoder is discarded and only the encoder part is used as a features extractor
- Like PCA, a compressed expression representing important features can be derived (can be used like non-linear PCA)
- It should be extracted as useful information that can determine the presence or absence of abnormal data or defects by using the reconstruction loss as an outlier score after learning only with normal data



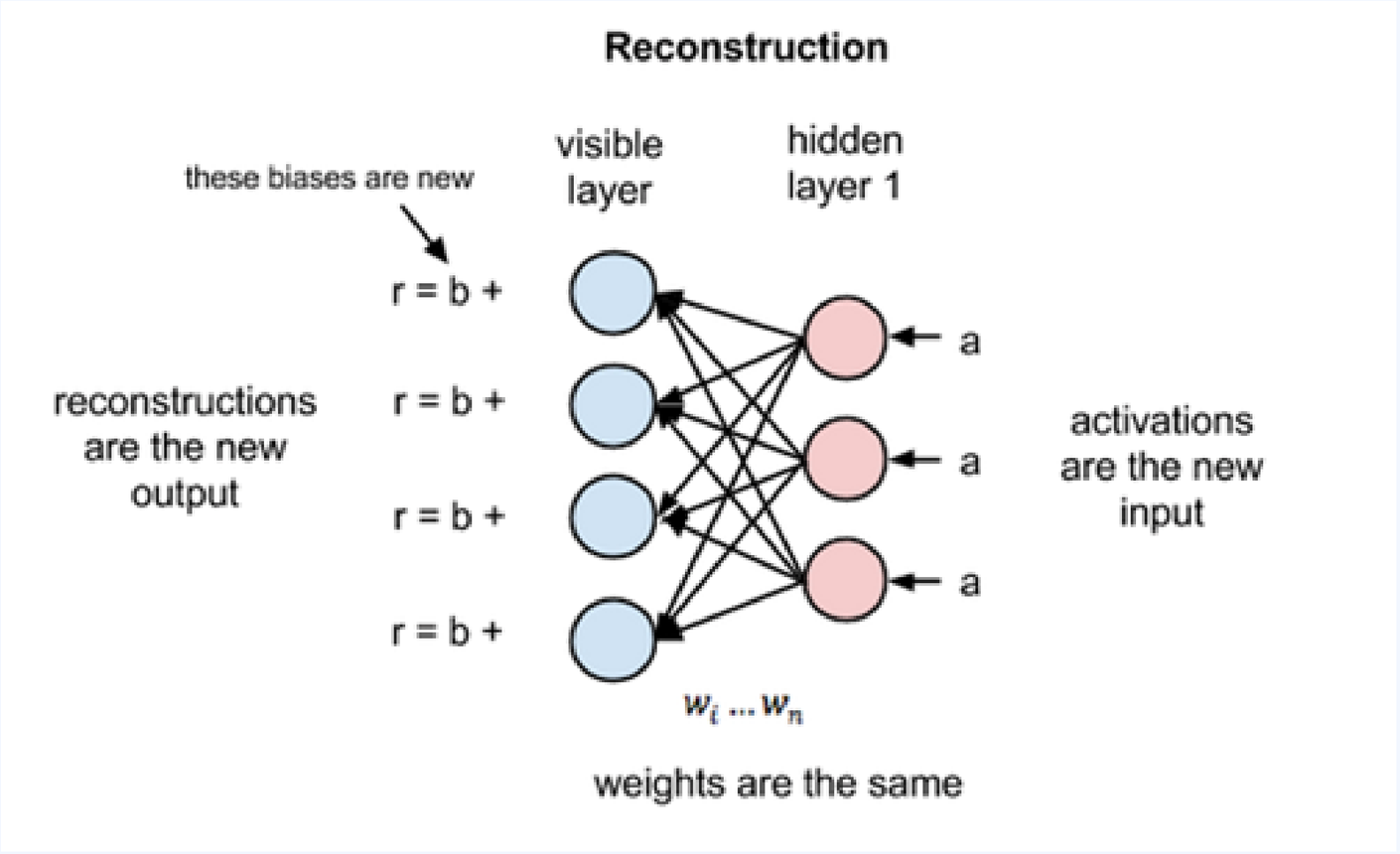
Paper Review

Problem of AutoEncoder



RBM

$$E(v,h) = -\sum_i \sum_j v_i W_{ij} h_j - \sum_i b_i v_i - \sum_j c_j h_j$$



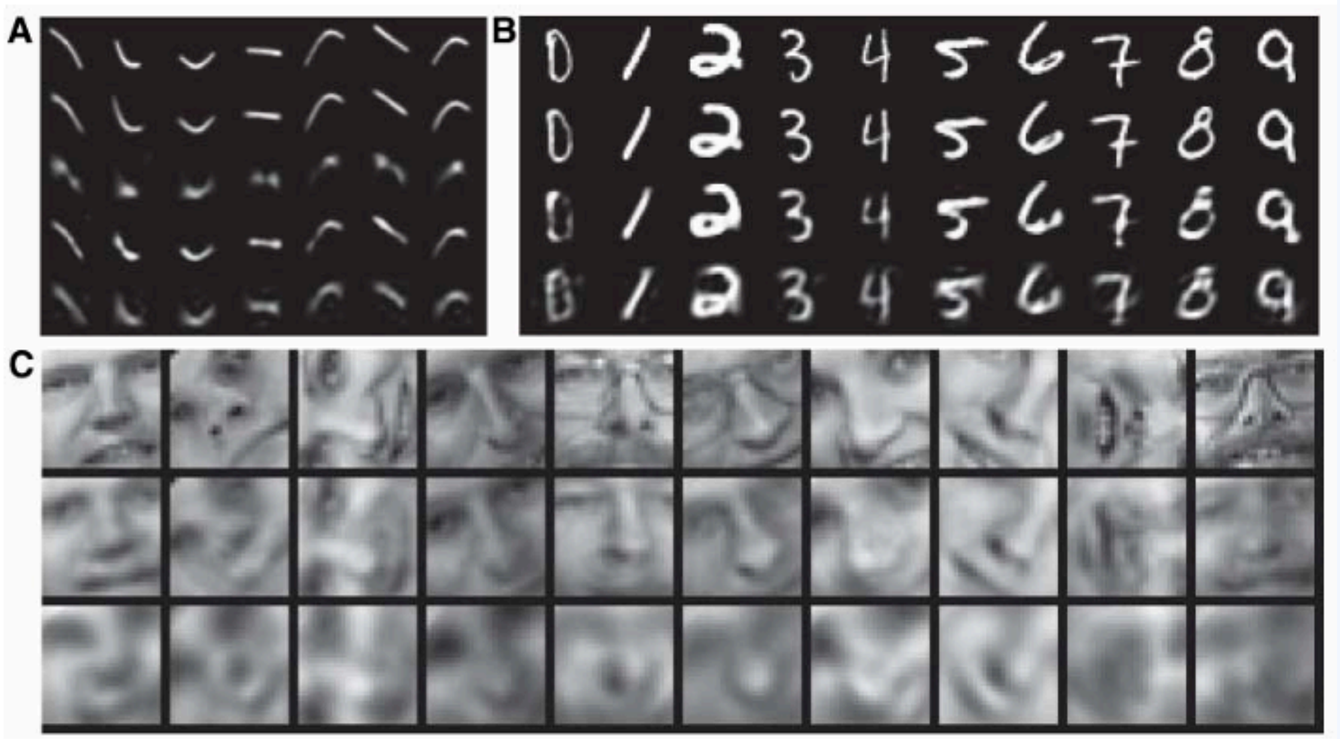


Fig. 2. (**A**) Top to bottom: Random samples of curves from the test data set; reconstructions produced by the six-dimensional deep autoencoder; reconstructions by "logistic PCA" (*8*) using six components; reconstructions by logistic PCA and standard PCA using 18 components. The average squared error per image for the last four rows is 1.44, 7.64, 2.45, 5.90. (**B**) Top to bottom: A random test image from each class; reconstructions by the 30-dimensional autoencoder; reconstructions by 30-dimensional logistic PCA and standard PCA. The average squared errors for the last three rows are 3.00, 8.01, and 13.87. (**C**) Top to bottom: Random samples from the test data set; reconstructions by the 30-dimensional autoencoder; reconstructions by 30-dimensional PCA. The average squared errors are 126 and 135.

Conclusion

AutoEncoder

- Features should be extracted as useful information
- Autoencoder can be applied in the following fields
    - Dimensionality reduction
    - Feature extraction
    - Anomaly detection
    - Image processing

THANK YOU