DL lab 7 -Autoencoders Answers

Link-https://github.com/Yashodh99/DL-LAB07

1. When above AE is used without activation functions, it is called a linear AE. Explain the relationship between linear AE and principal component analysis (PCA). Write the answer in a word file.

Autoencoders are neural networks that try to project the input data into a lower-dimensional space and then reproduce the input data as closely as possible. A Linear Autoencoder is an autoencoder where the nonlinear activation functions in hidden layers are discarded, so there is a linear relationship between the input and output. The connection between Linear Autoencoders and PCA in this case is more salient.

Principal Component Analysis (PCA) is a statistical method applied to the data for decreasing the dimensions while retaining as much information about its variance as possible. It does this by finding orthogonal axes- now called the principal components-capturing maximum variance of the data and projecting it onto them. PCA is majorly used for data compression and noise reduction.

When an autoencoder is completely linear, both the encoder and decoder functions are linear mappings. Then, such a linear autoencoder learns a latent space summarizing the most important features of the input data. It has been shown that, in case there exists a linear autoencoder, such a latent space is aligned with the principal components of data hence linear autoencoders are considered equivalent to PCA in terms of learned features.

Compared to PCA, in a linear autoencoder, the encoder serves to project the input data onto the low-dimensional principal component space, similar to the projection step in PCA. Then, the decoder projects this back linearly to approximate the input data, just as inverting the projection in PCA would do. Thus, while PCA uses matrix decomposition, a linear autoencoder achieves the same result through gradient-based learning methods.

1. Observe the model performance improvements between the above two models and give reasons for the observed improvements.

**Training vs. Validation Loss Trend:**

* In the first plot, while the training loss is monotonically decreasing, the trend of the validation loss is unpredictable since there are striking peaks and valleys.
* In the second graph, training and validation losses have a much smoother downward trend, suggesting better convergence.

**Stability of the Validation Loss:**

* In the first plot, the validation loss shows sharp fluctuations up and down, which may indicate instability in the training process.
* In this second plot, apart from some minor fluctuations in the validation loss, the general trend of loss is steadily decreasing and seems more stable in comparison with the first plot.

1. Observe the model performance improvements between the Image De-noising AE and the Vanilla CNN AE.
   * Explain the reasons for the observed improvements.

**Performance Improvements:**

* + Image Quality: The Denoising Autoencoder produces overall sharper images compared to the Vanilla CNN Autoencoder, especially in those cases when the input image is highly noisy. This fact is explained by the simple reason that it was exactly trained to reconstruct clean images from corrupted inputs.
  + Robustness to Noise: Since the denoising autoencoder is built for robustness against noise, this generally does better on test datasets having noisy inputs.
  + Smaller Loss: The specialized training objective could allow the denoising autoencoder to have smaller training loss compared to that of the vanilla CNN autoencoder.

**Reasons for Possible Improvement:**

* + Training Objective: While the objective of training for the denoising autoencoder is reconstruction from corrupted input, this forces it to focus more strongly on capturing the essential features of interest.
  + Data Augmentation: The addition of noise to the training examples can enhance the capability of generalization of the model. This, in turn, can make it more robust against variations, and this can improve performance on unseen data.
  + Feature Extraction: In general, the features extracted by a denoising autoencoder are more relevant, hence giving a better reconstruction, especially under challenging conditions.

1. Explain the differences between AE and Variational AE (VAE).

In AE, the goal is the minimum reconstruction error, for instance, the Mean Squared Error between input and output. Its focus is to learn the representation of data in compressed form. In VAE, this maximizes the data likelihood, adding a regularize to ensure that the latent space follows a certain distribution, generally Gaussian. This allows for sample generations.

In AE, the representation in latent space can be irregular and hence lead to clustering or overfitting, which would make the sampling of new data points difficult. In VAE, it forces the latent-space distribution to be continuous and structured so that it can easily sample and interpolate data points.