DL Lab 7 - Autoencoders

**2. Relationship Between Linear Autoencoder (AE) and Principal Component Analysis (PCA)**

Linear Autoencoder: A linear autoencoder is a neural network without activation functions, consisting of an encoder and a decoder. The encoder compresses the input data into a lower-dimensional latent space, and the decoder reconstructs the input from that compressed representation.  
  
Principal Component Analysis (PCA): PCA is a statistical technique that reduces data dimensionality by projecting it onto orthogonal axes that capture the most variance.  
  
Relationship:  
- Linear AEs and PCA both aim to reduce data dimensionality.  
- Linear AE approximates the same linear transformation as PCA, with the encoder learning a mapping akin to the principal components found by PCA.  
- While PCA solves an eigenvalue problem to directly find principal components, linear AE uses optimization techniques like gradient descent to approximate a similar projection.  
- In essence, a linear AE is a neural network-based approximation of PCA.

**4. Observed Model Performance Improvements: Dense AE vs. CNN AE**

Performance Improvements:  
- Better Feature Extraction: The CNN autoencoder performs better because CNN layers are more efficient at capturing spatial hierarchies in the image data, unlike dense layers, which treat all pixels equally.  
- Preserving Local Structure: CNN layers preserve the local structure (spatial relationships) of images, which is crucial for tasks like image reconstruction.  
- Parameter Efficiency: CNNs require fewer parameters to achieve similar or better performance compared to dense networks, leading to improved generalization and faster convergence.  
  
Reasons for Improvements:  
- CNN’s convolutional filters allow the model to learn more meaningful patterns such as edges and textures, making it more efficient at reconstructing the Fashion MNIST images.  
- The use of pooling layers in CNNs also allows for a more compact representation of the data without losing important spatial information.

**6. Observed Model Performance Improvements: Image Denoising AE vs. Vanilla CNN AE**

Performance Improvements:  
- Noise Robustness: The Image Denoising AE shows improvement because it is trained to reconstruct images from noisy inputs, making it more robust to noise and overfitting.  
- Generalization: By adding noise to the input data during training, the denoising autoencoder learns to generalize better, reducing overfitting that might occur with the Vanilla CNN AE.  
  
Reasons for Improvements:  
- The added noise forces the network to focus on reconstructing the essential features of the images, ignoring irrelevant details and improving generalization.  
- This technique acts as a form of regularization, making the model less likely to memorize the training data and instead learn meaningful patterns.

**7. Differences Between Autoencoder (AE) and Variational Autoencoder (VAE)**

Autoencoder (AE):  
- AEs are deterministic models that learn a mapping from input to latent space and back to input through an encoder and decoder.  
- The latent space is unconstrained, and the model does not impose any particular structure on it.  
- AEs focus on minimizing reconstruction loss, without any probabilistic interpretation.  
  
Variational Autoencoder (VAE):  
- VAEs introduce a probabilistic approach, learning a distribution over the latent space rather than a fixed mapping.  
- VAEs enforce a structure on the latent space by approximating it with a known distribution (usually Gaussian).  
- VAEs use a different loss function composed of two terms: the reconstruction loss and the Kullback-Leibler divergence, which measures how close the learned latent space distribution is to the prior distribution.  
- VAEs generate more realistic samples and are used in generative tasks because the structured latent space allows for meaningful interpolations between data points.