Auto Encoders

Eng. Ahmed Métwalli

December 24, 2024

Understanding Autoencoders

1. Basic Autoencoder Equations

Encoder Transformation:

$$z = f(x) = \sigma(W_e \cdot x + b_e)$$

where:

- x: Input data.
- W_e, b_e: Encoder weights and biases.
- σ: Activation function (e.g., ReLU).
- z: Latent-space representation.

Decoder Transformation:

$$x' = g(z) = \sigma(W_d \cdot z + b_d)$$

where:

- W_d, b_d: Decoder weights and biases.
- x': Reconstructed input.

Loss Function: Measures the difference between the input x and the reconstruction x'. Common choices include:

$$\text{Loss} = \frac{1}{n}\sum_{i=1}^{n}(x_i - x'_i)^2 \quad \text{(Mean Squared Error, MSE).}$$

2. Sparse Autoencoder

Modification: Adds sparsity constraint to the latent representation z, ensuring that most neurons are inactive.
Regularization:

$$\text{Loss} = \text{Reconstruction Loss} + \lambda \cdot \sum_{j=1}^{k}|z_j|$$

where:

- λ: Regularization coefficient.
- z_j: Latent-space activation.

3. Denoising Autoencoder

Modification: Corrupts input x with noise $\tilde{x}$ during training and reconstructs the clean input x.
Objective:

$$x' = g(f(\tilde{x})), \quad \text{where } \tilde{x} = x + \text{noise}.$$

Loss Function:

$$\text{Loss} = \frac{1}{n}\sum_{i=1}^{n}(x_i - x'_i)^2.$$

4. Variational Autoencoder (VAE)

Modification: Learns a probabilistic latent representation instead of a deterministic one.

Latent Space:

$$z \sim \mathcal{N}(\mu, \sigma^2)$$

where:

- $\mu = W_\mu \cdot x + b_\mu$: Mean of the latent distribution.
- $\sigma^2 = \exp(W_\sigma \cdot x + b_\sigma)$: Variance of the latent distribution.

Loss Function: Combines: 1. **Reconstruction Loss:**

$$\text{Reconstruction Loss} = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_i')^2.$$

2. **KL Divergence:**

$$D_{\text{KL}} = -\frac{1}{2}\sum_{j=1}^{k}\left(1 + \log\sigma_j^2 - \mu_j^2 - \sigma_j^2\right).$$

Final Loss:

$$\text{Loss} = \text{Reconstruction Loss} + D_{\text{KL}}.$$

5. Convolutional Autoencoder

Modification: Uses convolutional layers instead of dense layers, making it suitable for image data.

Equations:

$$z = \text{Conv2D}(x), \quad x' = \text{Conv2DTranspose}(z).$$

Loss Function: Measures the difference between the original image x and the reconstructed image x':

$$\text{Loss} = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_i')^2 \quad \text{(Mean Squared Error, MSE)}.$$

—

Comparison of Autoencoders

Type	**Algorithm**	**Advantages**	**Disadvantages**	**Evaluation Methods**
Basic Autoencoder	Encodes input into a latent space via an encoder. Decodes from latent space to reconstruct input. Optimizes reconstruction loss (e.g., MSE).	Simple and easy to implement. Good for dimensionality reduction. Non-linear feature extraction.	Poor generalization for complex data. Vulnerable to overfitting. Requires a good balance of latent space size.	Reconstruction Loss (e.g., MSE, MAE).
Sparse Autoencoder	Adds sparsity constraint to latent space (e.g., using $L1$-regularization or KL divergence). Encourages activation of only a few neurons.	Focuses on the most important features. Avoids trivial mappings. Better generalization.	Needs careful tuning of sparsity parameter. High computational cost due to regularization.	Reconstruction Loss. Sparsity Metric (e.g., activation statistics).
Denoising Autoencoder	Corrupts input with noise (e.g., Gaussian or masking noise). Trains to reconstruct clean input. Optimizes reconstruction loss.	Effective for denoising noisy data. Improves robustness to noise. Good for data augmentation.	Sensitive to the type and level of noise used. Requires sufficient training data for robustness.	Reconstruction Loss on noisy inputs. PSNR or SSIM (for image data).

Type	**Algorithm**	**Advantages**	**Disadvantages**	**Evaluation Methods**
Variational Autoencoder (VAE)	Maps input to a probability distribution in latent space. Adds KL divergence loss to match latent space to prior distribution (e.g., Gaussian). Decodes samples from latent space.	Useful for generative modeling. Captures meaningful latent structure. Generates new samples similar to input.	More complex optimization due to KL divergence. Trade-off between reconstruction quality and latent space regularity.	Evidence Lower Bound (ELBO). Reconstruction Loss. Quality of generated samples (visual inspection, FID score).
Contractive Autoencoder	Adds a penalty term to the loss to minimize the sensitivity of latent space to small input changes. Encourages robustness to perturbations.	Robust to small input variations. Good for extracting stable features.	High computational cost due to Jacobian matrix penalty. Limited applicability to large datasets.	Reconstruction Loss. Sensitivity Analysis.
Convolutional Autoencoder	Replaces dense layers with convolutional layers. Encodes spatial hierarchies in data. Optimizes reconstruction loss for image data.	Effective for image data. Captures spatial relationships. Fewer parameters compared to dense layers for large inputs.	Less suitable for non-image data. Requires careful architecture design to preserve spatial information.	Reconstruction Loss. PSNR, SSIM (for images).