Autoencoder

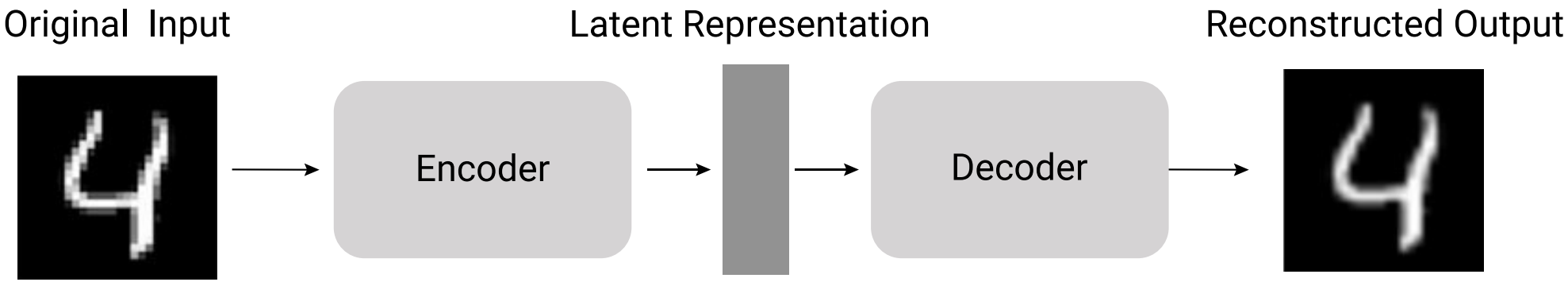
- Un autoencoder è un tipo di rete neurale il cui obiettivo è:
  - **Imparare una rappresentazione «compressa»** del dato iniziale
  - **Ricostruire** il dato originale
- Questi due task sono performati da:
  - **Encoder**
  - **Decoder**
- Addestramento: ridurre la differenza fra il dato originale e quello ricostruito

Autoencoder

- Applicazioni:
  - Dimensionality reduction
  - Anomaly Detection
  - Denoising
  - Data compression and generation

Autoencoder

- È un'architettura che effettua due fasi per
  - Ricostruire dati compressi
  - "Ripulire" dati da un eccessivo rumore
- I dati ricostruiti non sono identici ma *molto simili* agli originali



Autoencoder

- Un autoencoder è **data-specific**
  - Lavorano bene solo con dati molto simili a quelli su cui sono stati addestrati
- Un autoencoder è **lossy**
  - L'uscita sarà *sicuramente* diversa dall'input (anche solo per un bit)
- Un autoencoder apprende in maniera **non supervisionata**:
  - Non hanno bisogno di label esplicite per essere addestrati

Autoencoder

- Un autoencoder è formato da due reti distinte
  - Una per l'encoding: prende in input il dato e lo trasporta nello **spazio latente**
  - Una per il decoding: parte dallo spazio latente e ricostruisce il dato nello spazio di input
- Non devono necessariamente lavorare insieme
  - Es. Su una macchina abbiamo l'encoder, su un'altra il decoder

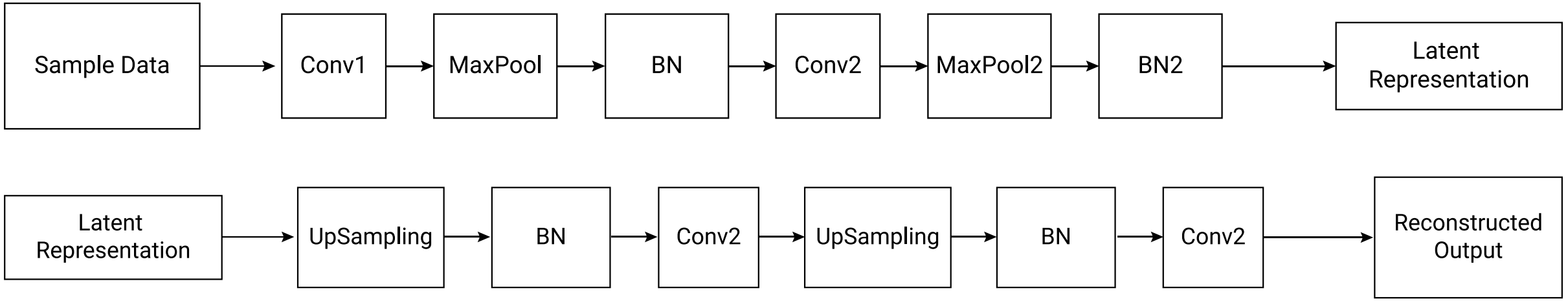
Autoencoder

- Obiettivo dell’addestramento:
  - Diminuire la distanza fra il dato originale e quello ricostruito
  - Funzione di costo: MSE
  - Backpropagation
- Idea chiave:
  - **Bottleneck** e compressione dell’informazione

Autoencoder - Varianti

- **Sparse** Autoencoder: imparano rappresentazioni **sparse** del dato in input
- **Denoising** Autoencoder: ricostruiscono dati «clean» da input rumoroso
- **Variational** Autoencoder: alla base di molte tecniche di generazione di dati
- **Contractive** Autoencoder: l'encoder è meno sensibile a piccole variazioni dell'input

Autoencoder Convolutivo



Autoencoder – Spazio latente

- Caratteristiche chiave:
  - **Dimensione ridotta**
  - **Presenza delle feature «importanti»**
  - **Rappresentazione Lossy**
- Cattura la vera «essenza» del dato in input in forma compatta