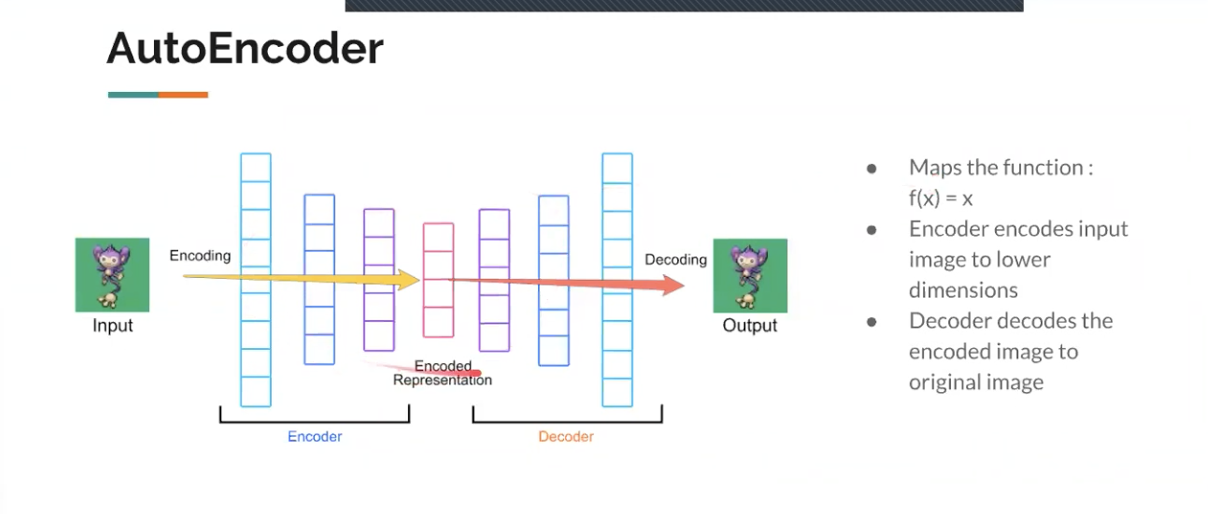
AutoEncoders Note

* Unsupervised neuralnets
* Encoder, bottleneck, decoder
* Learn representation that minimizes reconstruction loss
* Output = reconstructed input
* Fully connected neurons
* Learning weights by recirculation or backpropagation’
* Constraint the properties of the hidden layers
* If number of neuron in h < neurons in input layer → under complete, so network needs to learn sth
* Activation needs for decoder phase to perform non-linear version of PCA
* small latent layer and shallow network → underfit
* Larger hidden layer (at least size of input) + non-linearity to fit more complex
* Regularized auto-encoders, sparsity
* Sparsity



Progress and updates:

* Got recommendation engine to work for AutoEncoders, checked but still worried it might be wrong
* Fixed the constant validation loss/ RMSE issue by loading in and preprocessing the data exactly the same way as the original code
* Tried the recompile and train again but found that RMSE still plateaus at 0.85
* Tried more layers and more neurons in each layer but found that RMSE still plateaus at 0.86
* Tried getting more data but RAM crashed: 2M crashed, 5M crashed, 10M crashed → don’t think 20M will work either

Plan for tomorrow:

* Get more data, try training loop with multiple data
* <https://github.com/Melo96/SerreLab-multicue/blob/main/stage2/multi_input_parquet.py>
* Fine-tuning lambda, learning rate,
* More fine-tuning

Saturday Updates

* Change regularizer lambda to 0.01, results got worse, stuck at 0.92, both with 500 and 1000 epochs
* Layers = [512, 1024, 512], lambda = 0.002, test rmse = 0.86
* Layers = [512, 1024, 512], lambda = 0.002, test rmse = 0.88