

Artificial Neural Network

3rd Assignment – Shahid Beheshti University – Master’s Program

“December 11, 2023”

Hello everyone, I hope you’re doing well. This is the third assignment for the Artificial Neural Networks course. The deadline for this assignment is **December 29, 2023**. All students are expected to submit their homework on time. feel free to ask any questions regarding the exercises. Please noted that for the third part of your assignment, you are expected to write a detailed report.

Part 1 (Autoencoders)

- a) Explain what problems sampling from the $q(z|x)$ distribution leads to in Variational-Autoencoders, explain how the sampling process can be rewritten to solve the problems, and give an explanation of how this trick works.
- b) How can you use a VAE to measure the degree of imbalance in a dataset? What are the advantages and disadvantages of this method compared to other techniques?
- c) How does the hyperparameter beta affect the trade-off between reconstruction accuracy and disentanglement in a variational autoencoder?

Part 2 (Recurrent neural networks)

- a) Design and describe a Vanilla RNN model to solve an anomaly detection task in time series. Discuss potential pitfalls the model might face.
- b) Given a sequence of inputs, draw a detailed diagram showing the flow of information through a GRU cell at two adjacent time steps.
- c) Analyze the impact of “peephole connections” in LSTMs. How do they theoretically improve the LSTM’s ability to maintain its cell state? Include a discussion about the situations where peephole connections may not be beneficial.

Part 3

This part involves developing an encoder-decoder based architecture to perform image captioning on the Flickr30K dataset. The Flickr30K dataset contains 31,000 images collected from Flickr, together with 5 reference sentences provided by human annotators.

- Construct a baseline image captioning model employing a pre-trained CNN (e.g., VGG16) for image feature extraction and a simple RNN for caption generation. Do not forget to use the previously discussed good practices in training an artificial neural network.
- What scores did use in the previous part? Explain BLUE, METEOR, CIDEr and ROUGE scores and evaluate baseline performance using BLEU and ROUGE scores to set a benchmark for subsequent improvements.
- To address the limitations of the baseline RNN's capacity for long-term dependencies, upgrade the decoder network to an LSTM. Measure any performance improvements against the baseline model using the same evaluation metrics.
- Enhance the LSTM-based model by incorporating a soft attention mechanism, allowing the model to focus on different regions of an image when generating each word in the caption. Compare performance improvements over previous models both quantitatively and qualitatively.
- Bonus: Implement pre-trained word embeddings such as GloVe or Word2Vec within the decoder to enrich the semantic understanding of captions. Evaluate the impact of these pre-trained embeddings on the model's ability to generate coherent and contextually relevant captions, especially considering the limited vocabulary size of the Flickr30K dataset.



Figure 1 Data Example