Generative ML for time series

Johan Vik Mathisen

April 4, 2024

0.1 Stage 1

As mentioned in the section on representation learning, one needs to determine a set of tasks one wish to evaluate on, in order to say anything about the quality of the representations. We evaluate the representations based on two tasks

0.1.1 Evaluation metrics

- **Reconstruction**: We evaluate the models ability to reconstruct the original data from latent representation. Success indicating perservation of information.
- **Downstream classification**: We evaluate the latent representations on its ability linear classification.
- **Training time**
- **Number of parameters**
-

0.1.2 Reconstruction

0.1.3 Classification

0.1.4 Codebook investigations

In the two tokenization models, how does the codebooks differ? Look at codebook utlization. Histograms across dimensions?

0.2 Stage 2

0.2.1 Evaluation metrics

- **IS**:
- **FID**:
- **Visual inspection**:
- **Token usage**:
- **Generating distribution**:

0.3 Ablation studies

TODO: Could it be interesting to investigate different masking designs for time series? Is cosine the ubiquitous choice

The masking design does not affect (gradients) training, only the iterative decoding. Can do tests on trained models!

Different sampling scheme? Instead of TopK try Nucleus sampling?