CERN
openlab

Conditional VAE in Healthcare

Author: Sara Zoccheddu

Supervisor(s): Nicola Serra, Olivia Jullian Parra

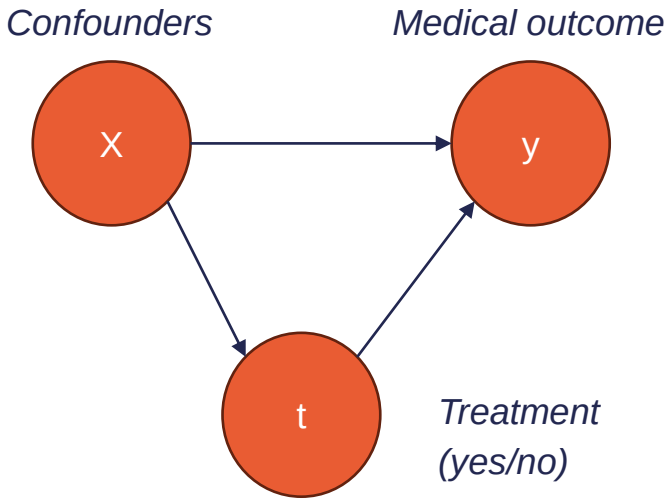
Summary

- General context: Machine Learning for Causal Inference in Multimorbidity
- Conditional Variational Autoencoder
- Task: Reconstruction of two gaussian distributions using CVAE
- Conditional VAE applied to the IST dataset
- Results

Machine Learning for Causal Inference in Multimorbidity

Multimorbidity: coexistence of two or more chronic conditions in the same individual

Conceptual model

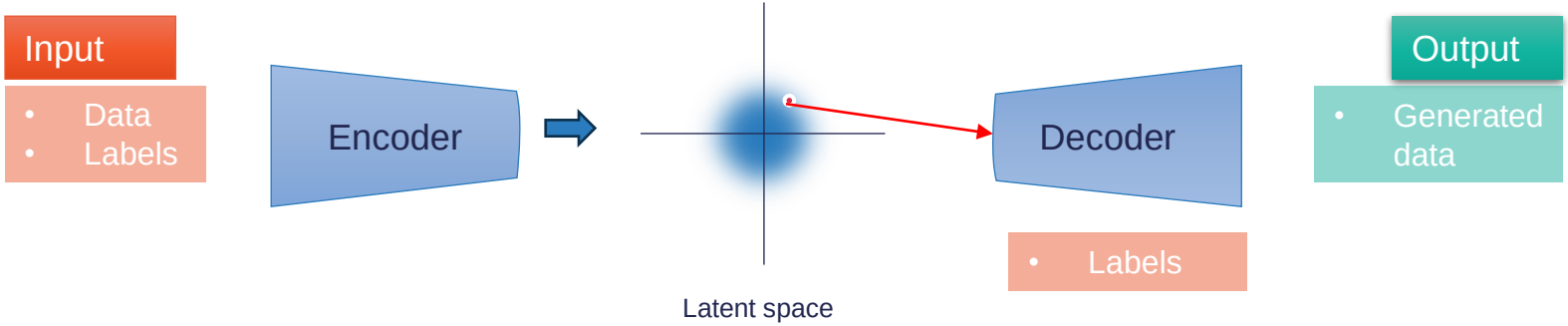


- **X:** patient's personal information (age, BMI, blood pressure, other previous treatments, etc…)
- **t:** if the treatment (i.e., aspirin) was taken or not
- **y:** outcom (i.e., dead, not recovered, recovered…)

We cannot ignore the link between X and t: patient's confounders affect the treatment.

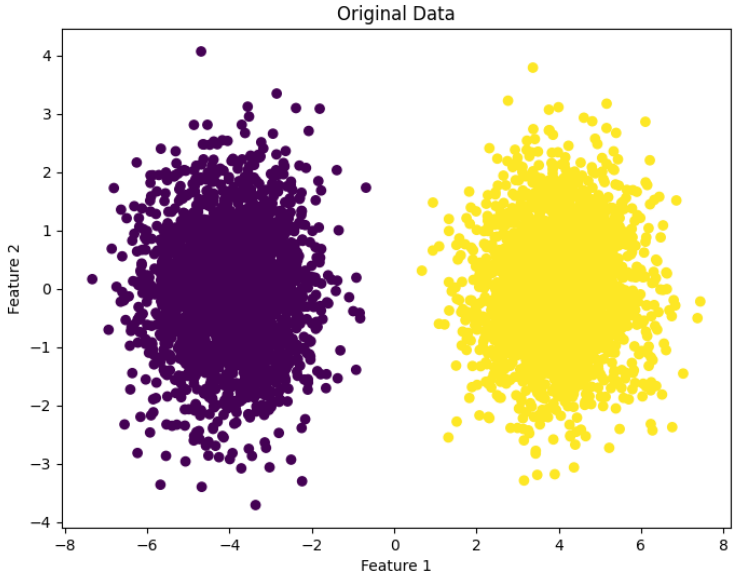
Conditional Variational Autoencoder (CVAE)

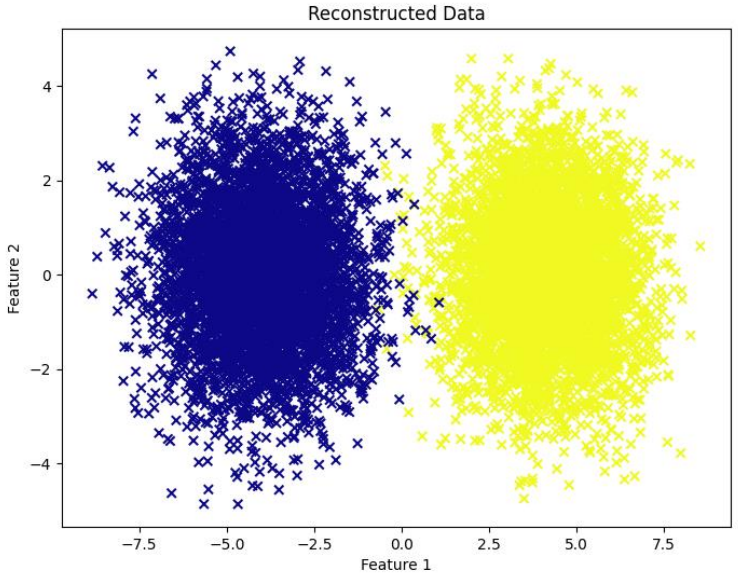
- A **variational autoencoder** (VAE) is a **generative model**: it should be able to generate fake samples that look like original samples from our data.
- Data is passed to the **encoder**.
- The encoder returns a **probability distribution** in the latent space.
- One point is sampled from the distribution and it is passed to the **decoder.**
- In a **Conditional VAE**, also **labels** are passed both to the encoder and the decoder.



Task: **Reconstruction of two gaussian distributions using CVAE**

Goal: Learn how to hard-code a CVAE and apply it to a bidimensional dataset.





Conditional VAE applied to the IST dataset

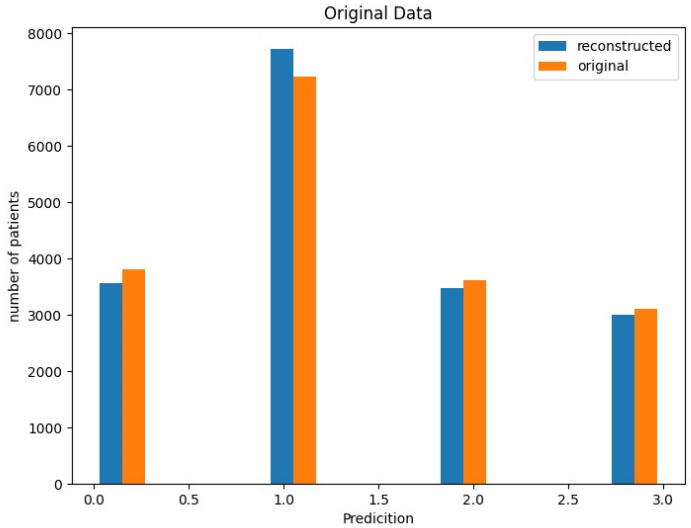
- The International Stroke Trial (IST) was conducted between 1991 and 1996.
- RCT (Randomized Controlled Trial): very low bias.
- **Goal**: understand whether early administration of aspirin, heparin, both or neither influenced the clinical course of acute ischemic stroke.
- **Idea**: apply the previous CVAE model to the IST dataset and see whether results are consistent with the ones obtained with the RCT.

❖ https://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-12-101

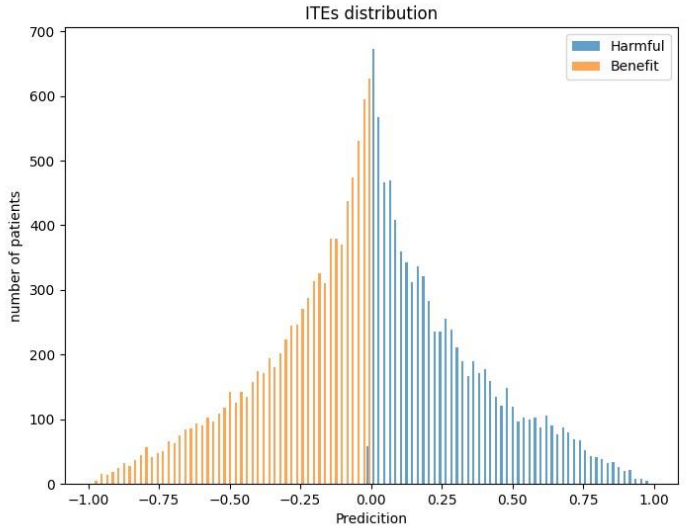
"The International Stroke Trial database" - Peter AG Sandercock, Maciej Niewada, Anna Członkowska and for the International Stroke Trial Collaborative Group

Conditional VAE applied to the IST dataset

Step 1: **Reconstruct the original data with the CVAE model**

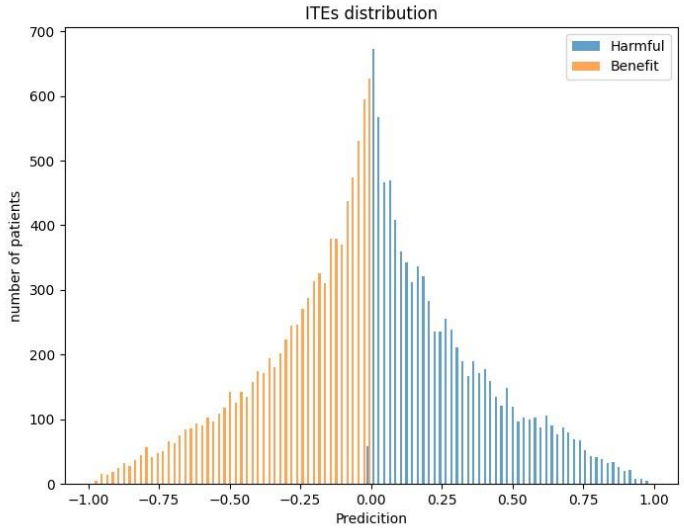


Step 2: **Study the effect of administration of Heparin**



Conditional VAE applied to the IST dataset

Step 2: **Study the effect of administration of Heparin**



- **ITE** (Individual Treatment Effect):
  - Difference between probability of dying or not recovering by taking Heparing or not taking Heparin
  - If ITE > 0: Treatment is **harmful**
  - If ITE < 0: Treatment is **beneficial**
- Conclusion: The effect of Heparin is meaningless on a large scale.

Comparison of results between RCT an CVAE

- The original data was well reconstructed by the CVAE model.
- The analysis of data from the RCT and the analysis of data generated by the CVAE gave similar results.

Why would you use a CVAE instead of a RCT in Healthcare?

- To have good results in RCT you would need a huge number of patients, therefore RCTs are not always feasibile.
- CVAE allows to generate new data.
- CVAE allows to model causality taking into consideration all the variables representing a patient.
- Useful for studying multimorbidity and cases in which more than one treatment is involved.

Thank you!

CERN
openlab