

Practice exercises 7

You will learn how to adapt transformers (i.e. HuggingFace BERT) to build multimodal transformers that combine text with categorical and numerical features for a binary classification task. Detailed instructions and explanations are provided in the accompanying Jupyter notebook and the additional resource files.

You will implement the following [combining](#) strategies:

7.1 concat: concatenate transformer output, numerical features, and categorical features all at once before the final classifier.

7.2 individual_mlps_on_cat_and_numerical_feats_then_concat: separate MLPs on categorical features and numerical features, then the concatenation of transformer output, with processed numerical features, and processed categorical features before the final classifier.

7.3 attention_on_cat_and_numerical_feats: attention-based summation of transformer outputs, numerical features, and categorical features queried by transformer outputs before the final classifier.

Note: The notebook serves as the main file for the practice exercise and contains the data preparation and model configuration steps. To get started with the exercise you need to upload the multimodal transformers.zip into your current working session (i.e. Gnote) and unzip it. Instructions for unzipping are provided in the notebook. Then, follow the instructions in the notebook in order to complete the task.