TVAESynthesizer

enforce_min_max_values: Control whether the synthetic data should adhere to the same min/max boundaries set by the real data

enforce_rounding: Control whether the synthetic data should have the same number of decimal digits as the real data

locales: A list of locale strings. Any PII columns will correspond to the locales that you provide.

epochs: Number of times to train the VAE. Each new epoch can improve the model.

cuda: Whether to use CUDA, a parallel computing platform that allows you to speed up modeling time using the GPU

batch_size: Number of data samples to process in each step. This value must be even, and it must be divisible by the `pac` parameter (see below). Defaults to `500`.

compress_dims: Size of each hidden layer in the encoder. Defaults to `(128, 128)`.

decompress_dims: Size of each hidden layer in the decoder. Defaults to `(128, 128)`.

embedding_dim: Size of the random sample passed to the Generator. Defaults to `128`.

l2scale: Regularization term. Defaults to `1e-5`.

loss_factor: Multiplier for the reconstruction error. Defaults to `2`.

CTGANSynthesizer

batch_size: Number of data samples to process in each step. This value must be even, and it must be divisible by the `pac` parameter (see below). Defaults to `500`.

discriminator_dim: Size of the output samples for each one of the Discriminator Layers. A Linear Layer will be created for each one of the values provided. Defaults to `(256, 256)`.

discriminator_decay: Discriminator weight decay for the Adam Optimizer. Defaults to `1e-6`.

discriminator_lr: Learning rate for the discriminator. Defaults to `2e-4`.

discriminator_steps: Number of discriminator updates to do for each generator update. Default `1` to match the original CTGAN implementation

embedding_dim: Size of the random sample passed to the Generator. (Default `128`)

generator_decay: Generator weight decay for the Adam Optimizer. Defaults to `1e-6`

generator_dim: Size of the output samples for each one of the Residuals. A Residual Layer will be created for each one of the values provided. Defaults to `(256, 256)`.

generator_lr: Learning rate for the generator. Defaults to `2e-4`.

log_frequency: Whether to use log frequency of categorical levels in conditional sampling. Defaults to `True`.

pac: Number of samples to group together when applying the discriminator. Defaults to 10.

enforce_min_max_values: Control whether the synthetic data should adhere to the same min/max boundaries set by the real data

enforce_rounding: Control whether the synthetic data should have the same number of decimal digits as the real data

epochs: Number of times to train the GAN. Each new epoch can improve the model.

verbose: Control whether to print out the results of each epoch. You can use this to track the training time as well as the improvements per epoch.