Revisiting Deep Learning Models for Tabular Data

I. FT_Transformer:

a. Architecture

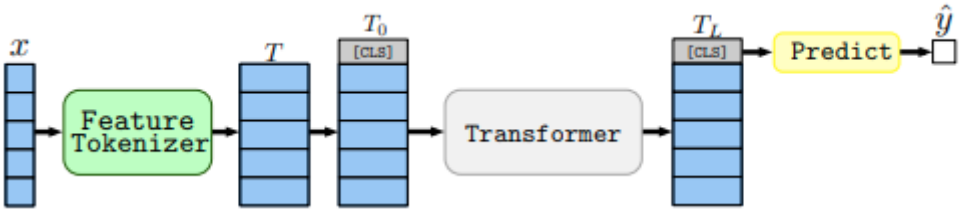
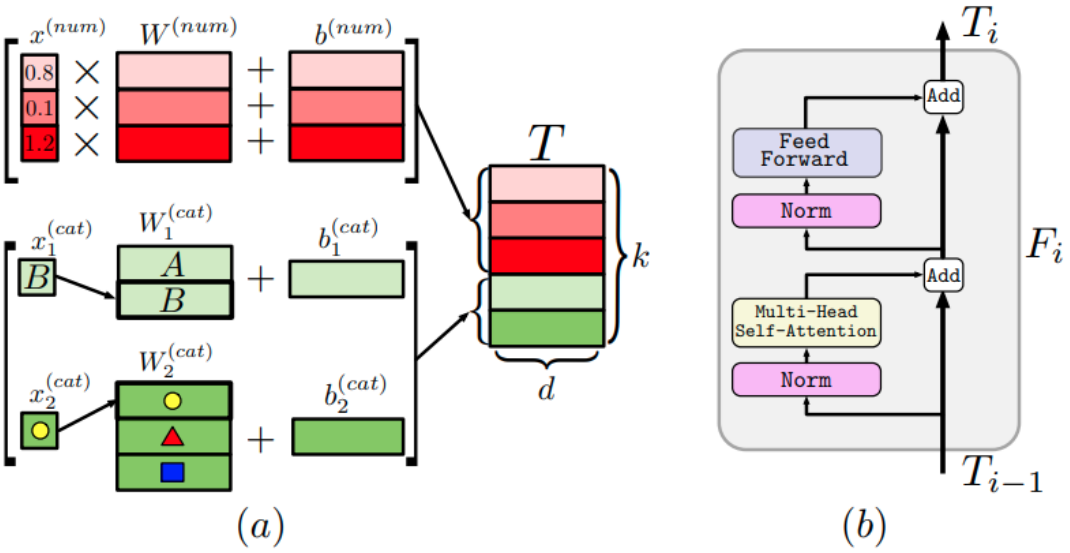


Figure 1: Architecture of FT-transformer.

b. Explanation:

- Feature Tokenizer transforms the input features x to embeddings $T \in \mathbb{R}^{k \times d}$, f_j: lookup table $W_j \in \mathbb{R}^{S_j \times d}$
- x^{num}: continue feature:
  $T_j^{num} = b_j^{num} + x_j^{num} \cdot W_j^{num}$
- x^{cat}: category feature:
  $T_j^{cat} = b_j^{cat} + e_j^T W_J^{cat}$
- $T = stack[T_1^{num}, , T_{k^{num}}^{num}, T_1^{cat}, , T_{k^{cat}}^{cat}] \in \mathbb{R}^{k \times d}$



a.Feature Tokenizer for numerical and two categorical features
b. One Transformer layer.

c. Prediction:

$$\hat{y} = \texttt{Linear}(\texttt{ReLU}(\texttt{LayerNorm}(T_L^{\texttt{[CLS]}}))).$$

II. Dataset

https://www.kaggle.com/datasets/shrutimechlearn/churn-modelling

CODE: https://github.com/HangBich/Machine-Learning/blob/main/rtdl.py

III. Results

	ACCURACY	ROC_AUC_SCORE
TRAIN	0.7945	0.5
TEST	0.8035	0.5