

딥러닝 모델 추천

tabnet

딥러닝의 한계

- features constructed from tabular data are often correlated, so a small subset of features are responsible for most of the predictive power,
- missing data in the form of NULL values in a database, and
- there is usually a strong class imbalance in the labels (in the supervised setting), for instance, users prefer only a small collection of movies in a movie catalogue.

딥러닝의 한계

- So, XGBoost, CatBoost, LightGBM is crowd favorites.
- Decision tree-based algorithms, Gradient boosted trees.

딥러닝의 한계

- 더욱이, 예측의 해석을 주는 요인을 찾지 못한다.

Tabnet

- Google research이며, 2019년 8월에 처음 소개되었다.
- 더 발전된 버전으로 20년 2월에 다시 발표

Tabnet: Keywords

- Interpretable deep learning
- Tabular data
- Learnable mask, Attention
- self-supervised

Soft feature selection

- 다른 트리 모델은 hard feature selection을 사용(아마 모든 특성 적용)
- Soft feature selection은 masking을 통해 이루어지는데 masking이란 feature에 mask를 씌운다고 생각하자.
- 이 때 mask는 sparse matrix가 되어야한다.

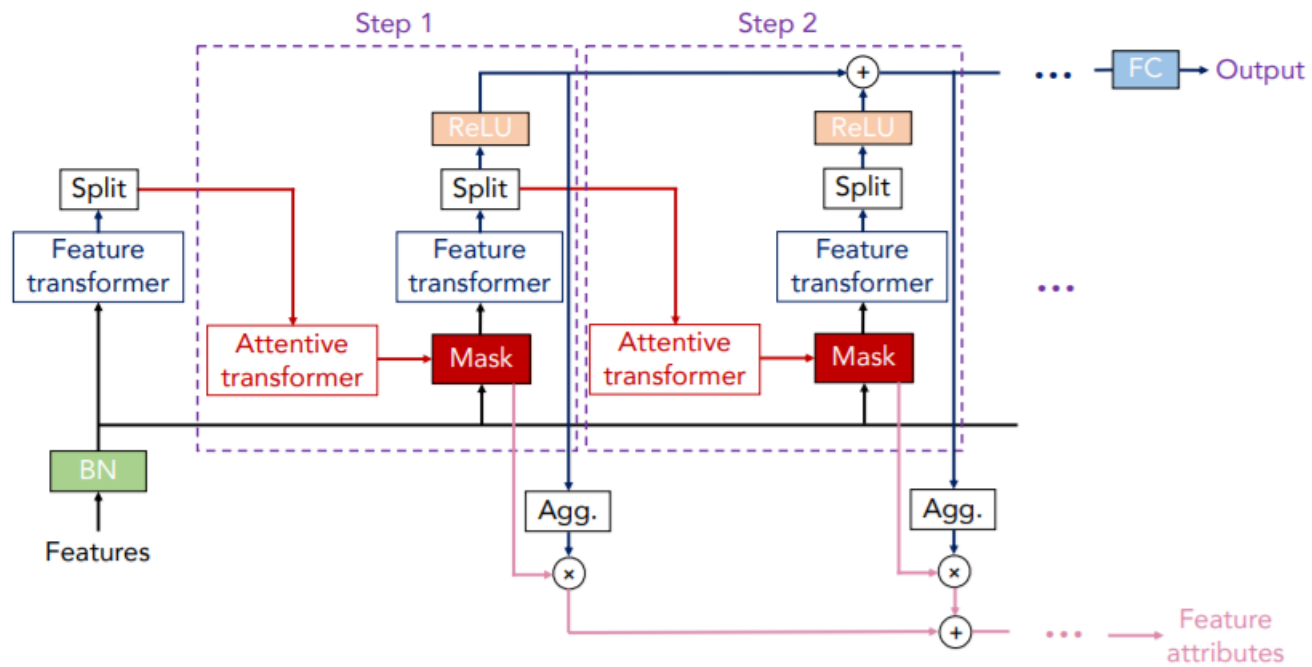
Integration with DNN and Decision tree

- Representation capacity (Deep learning)
 - Feature selection property (Tree model)
- > tree-based model

Integration with DNN and Decision tree

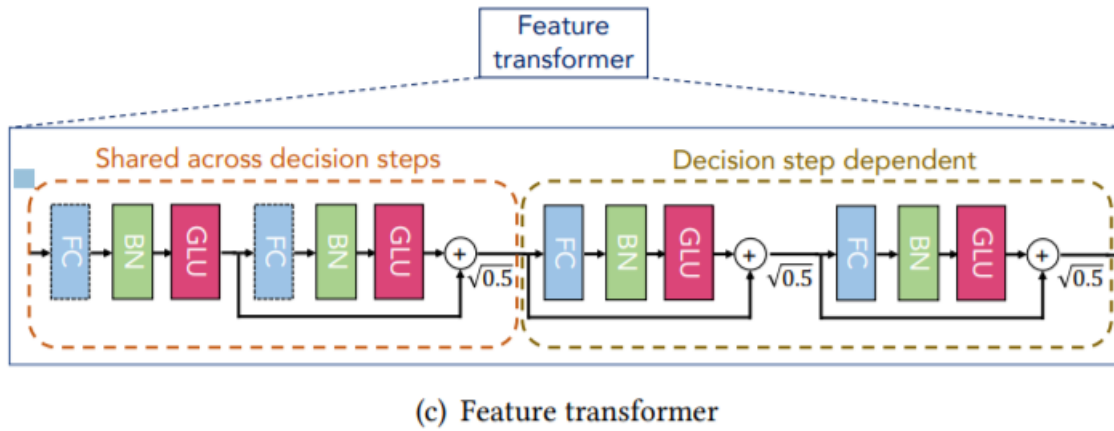
- Sequential attention을 통해 controllable label sparsity를 통한 soft feature selection ability를 갖게 된다.

Tabnet



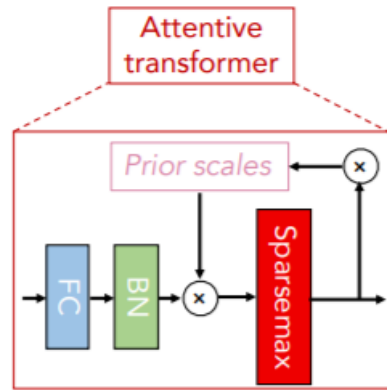
(a) TabNet encoder architecture

Tabnet



the feature transformer: where the feature processing and engineering happens. There are two shared feature blocks to allow processed features to be shared between decision steps, and two decision step-dependent feature processors.

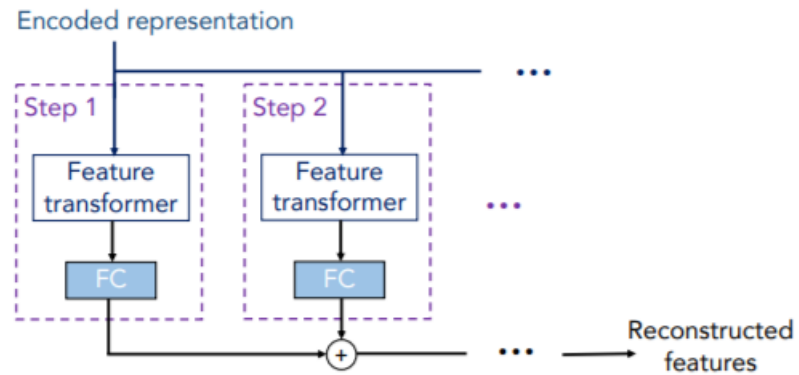
Tabnet



(d) Attentive transformer

the attentive transformer: where the learning of the sparse mask takes place. A sparsemax layer⁶ is used to promote sparsity in the learnable mask while it learns to attend to certain features. Feeding back information of *prior scales* allows TabNet to learn and control how much a feature has been used before in prior decision steps. A hyperparameter can be specified beforehand to control the amount of feature reuse between decision steps.

Tabnet



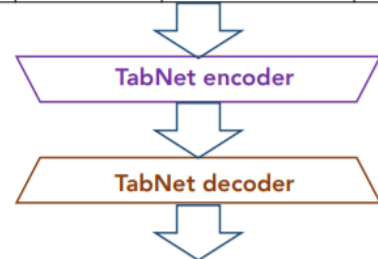
(b) TabNet decoder architecture

Tabnet decoder architecture

Tabnet: semi-supervised learning

Unsupervised pre-training

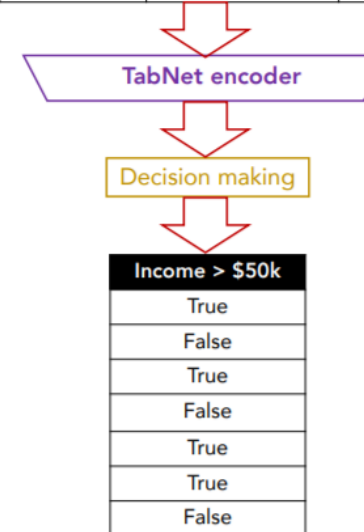
Age	Cap. gain	Education	Occupation	Gender	Relationship
53	200000	?	Exec-managerial	F	Wife
19	0	?	Farming-fishing	M	?
?	5000	Doctorate	Prof-specialty	M	Husband
25	?	?	Handlers-cleaners	F	Wife
59	300000	Bachelors	?	?	Husband
33	0	Bachelors	?	F	?
?	0	High-school	Armed-Forces	?	Husband



Age	Cap. gain	Education	Occupation	Gender	Relationship
		Masters			
		High-school			Unmarried
43					
	0	High-school		F	
			Exec-managerial	M	
			Adm-clerical		Wife
39				M	

Supervised fine-tuning

Age	Cap. gain	Education	Occupation	Gender	Relationship
60	200000	Bachelors	Exec-managerial	M	Husband
23	0	High-school	Farming-fishing	M	Unmarried
45	5000	Doctorate	Prof-specialty	M	Husband
23	0	High-school	Handlers-cleaners	F	Wife
56	300000	Bachelors	Exec-managerial	M	Husband
38	10000	Bachelors	Prof-specialty	F	Wife
23	0	High-school	Armed-Forces	M	Husband



Income > \$50k
True
False
True
False
True
True
False