

Tabnet系列讲解

【核心架构介绍】

TabNet系列讲解

【核心架构介绍】

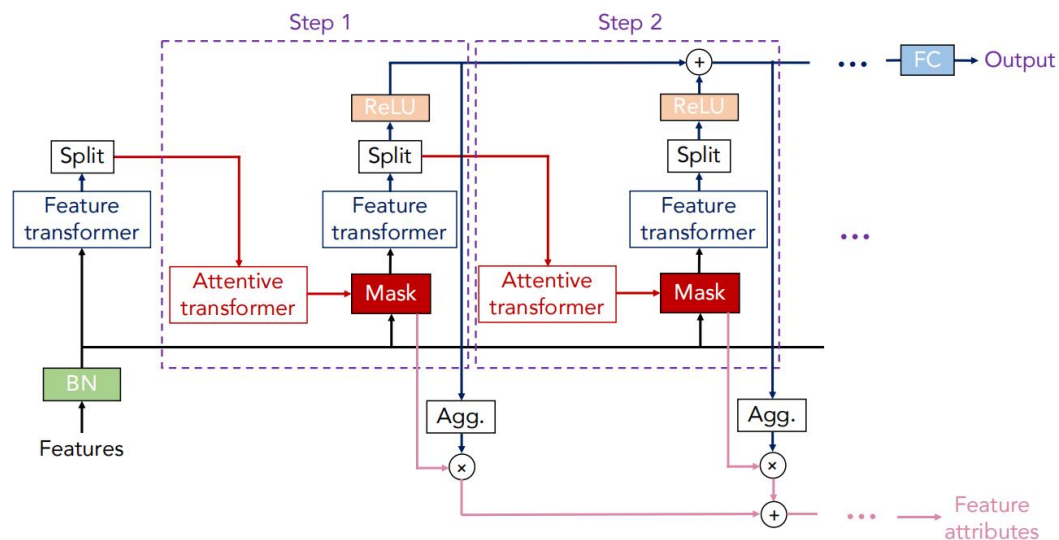
回顾-TabNet

- 背景：表格数据的深度学习仍未得到充分探索，集成决策树(DTs)的变体仍然主导着大多数应用。
- TabNet：采用DNN架构实现树模型的决策流的深度学习模型，使其适应于表格数据的分类、回归任务，并具有较好的可解释性。

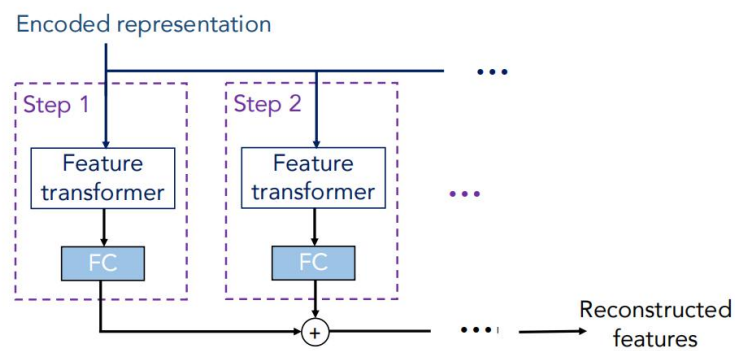
目录

- 决策流形的实现
- 可解释性的实现
- 自监督学习

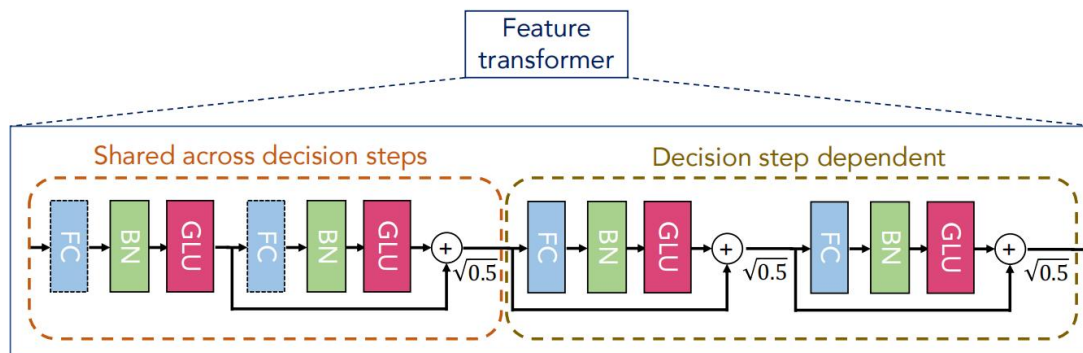
TabNet核心架构



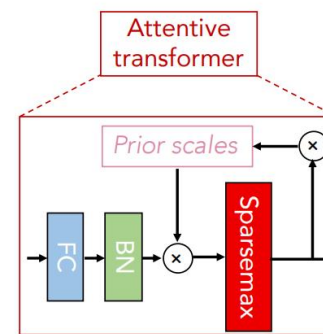
(a) TabNet encoder architecture



(b) TabNet decoder architecture

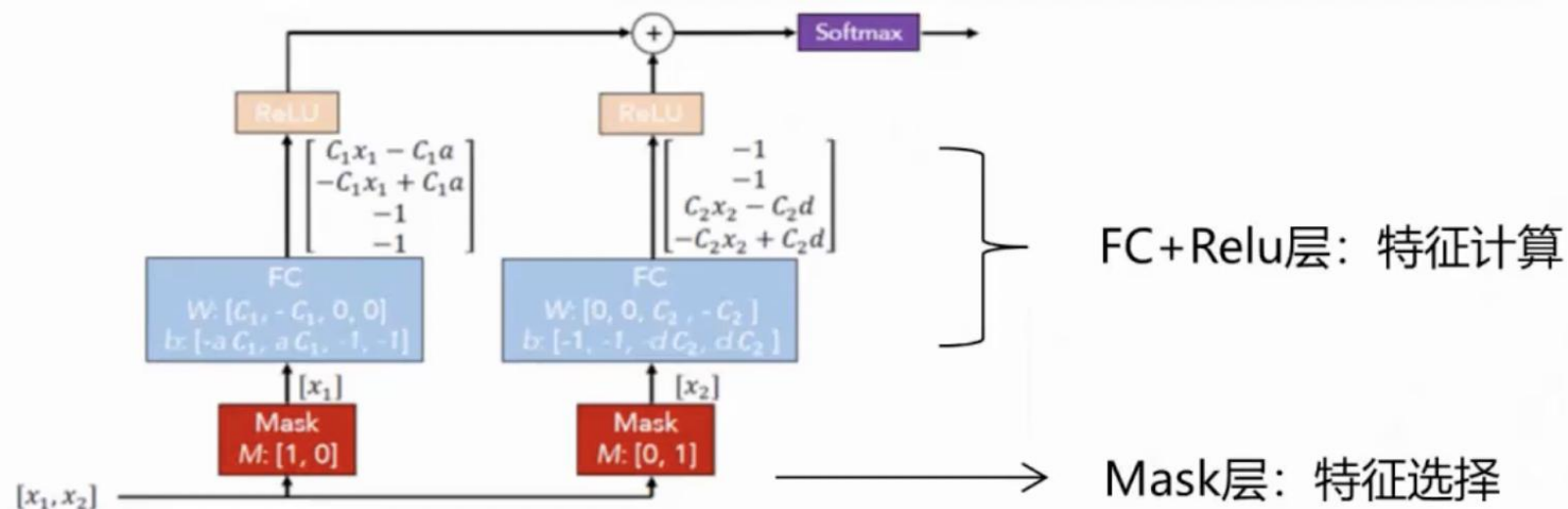
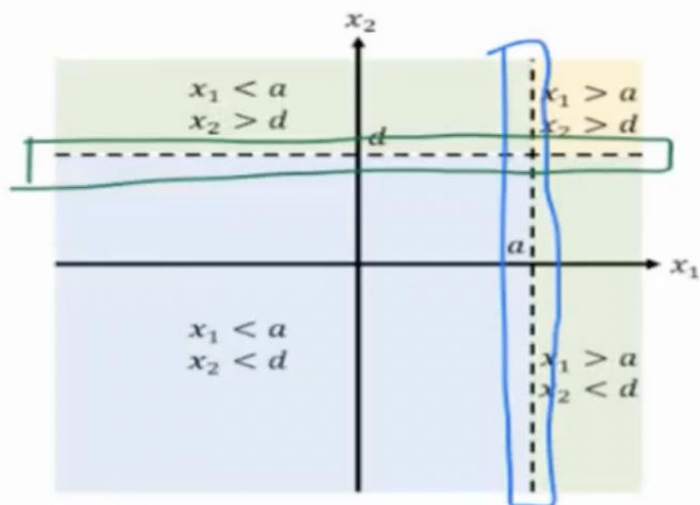


(c)



(d)

决策流形的实现 (demo版)



以特征 x_1 的条件判断为例:

$$\begin{bmatrix} C_1x_1 - C_1a \\ -C_1x_1 + C_1a \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} f(x_1) \\ -f(x_1) \\ -1 \\ -1 \end{bmatrix}$$

若 $f(x_1) > 0$, 则Relu后

$$\begin{bmatrix} \text{relu}(f(x_1)) \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

若 $f(x_1) < 0$, 则Relu后

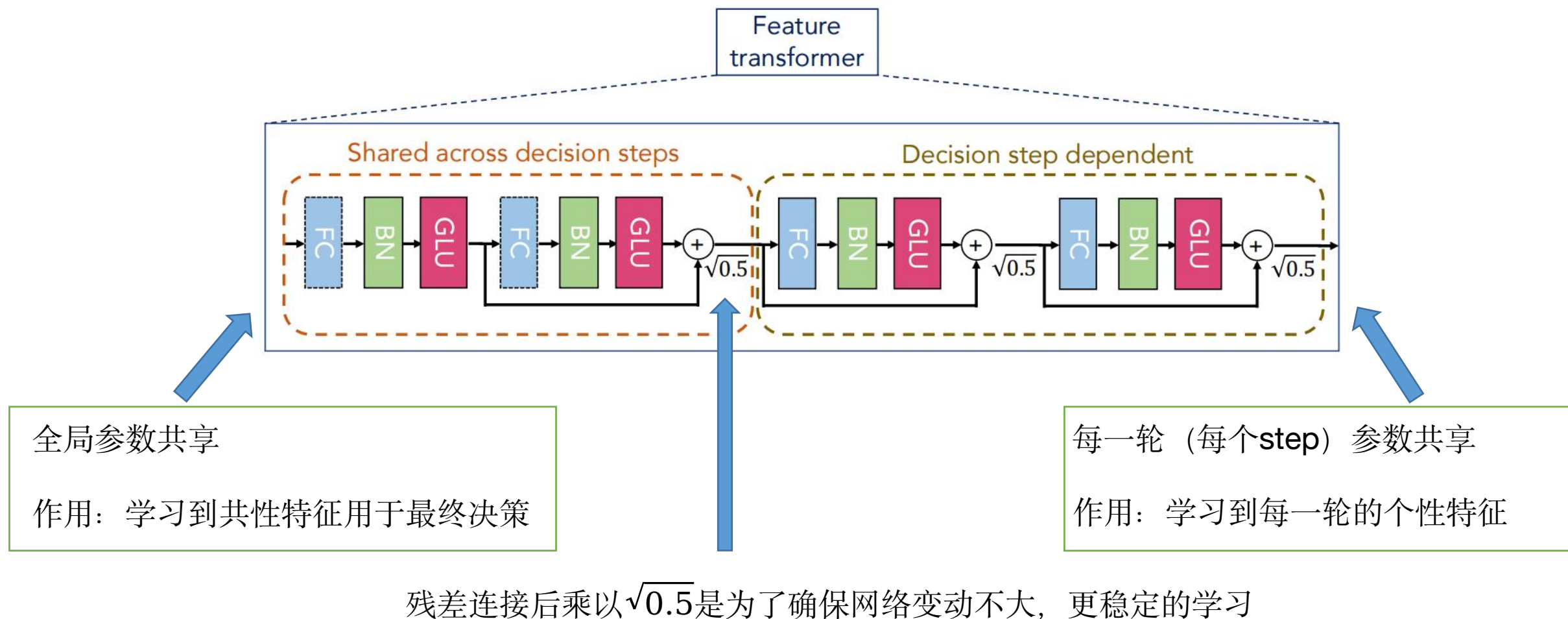
$$\begin{bmatrix} 0 \\ \text{relu}(-f(x_1)) \\ 0 \\ 0 \end{bmatrix}$$

对应树的决策流形

$$\begin{bmatrix} x_1 > a \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ x_1 < a \\ 0 \\ 0 \end{bmatrix}$$

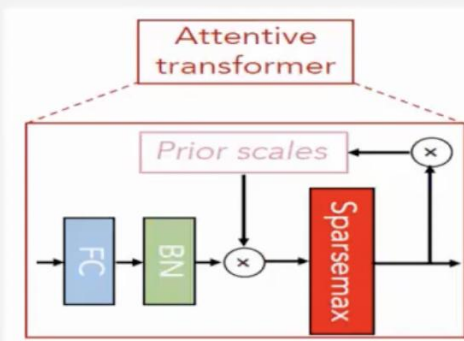
特征计算



Mask:Attentive Transformer (特征选择)

Attentive Transformer: 学习一个“软”的Mask，用于选择重要特征，让每轮决策下的学习不再浪费在无关特征上，因此模型变得更有效率。

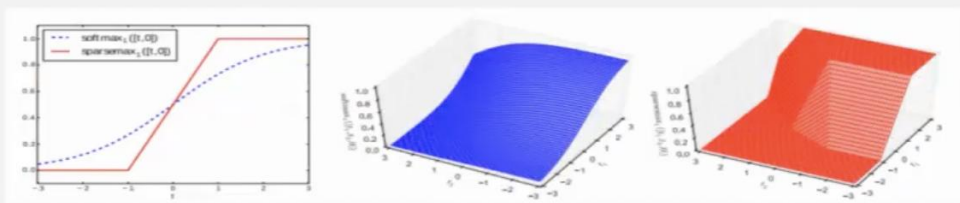
$$M[i] = \text{sparsemax}(P[i-1] \cdot h_i(a[i-1]))$$



其中,

- $a[i-1]$ 是上一轮决策中Split层划分出来的输出,
- $h_i(\cdot)$ 代表FC+BN层,
- $P[i]$ 是先验尺度项 (Prior Scale Term), 用于告知模型某个特征在历史决策里被使用的程度。

特征选择: $M[i] \cdot \text{feature}$



sparsemax使得特征挑选也变得稀疏化

先验尺度项:
$$P[i] = \prod_{j=1}^i (\gamma - M[j])$$

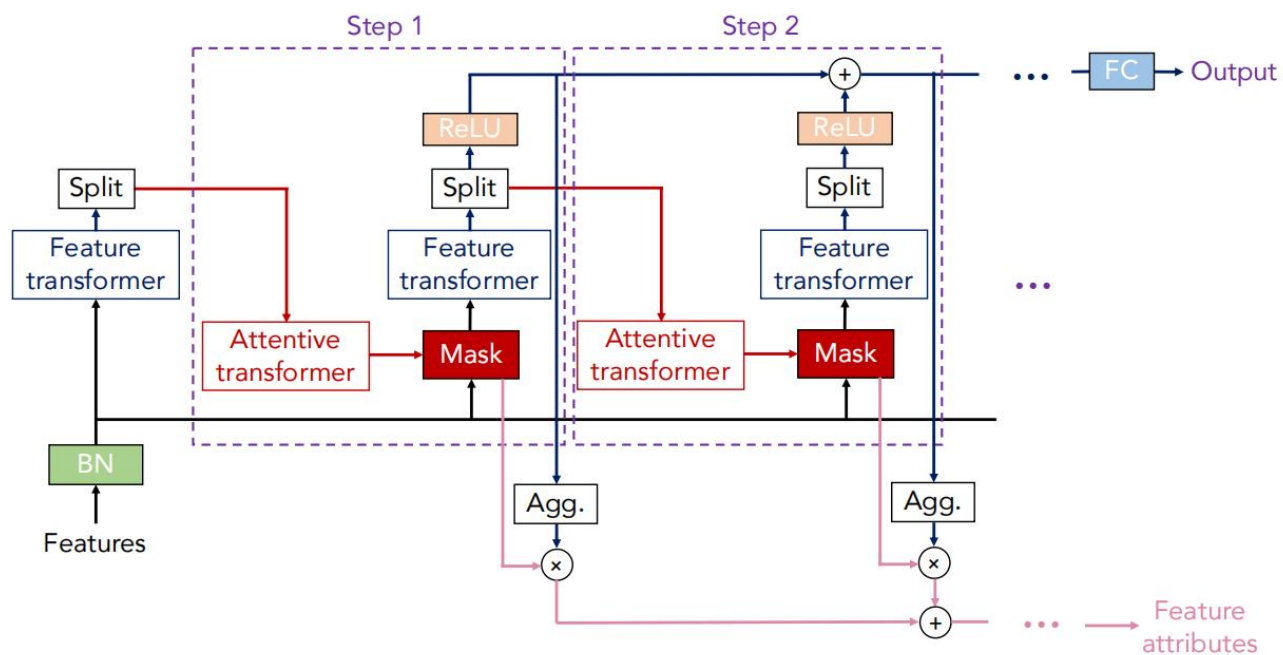
i 为第几轮决策, 其中 γ 是个松弛参数, 当 $\gamma=1$ 时, 一个特征只允许被使用一次, γ 增加, 一个特征可在多个决策轮数下被灵活使用。

另外, Batch下的每个样本都有一个 (特征注意力权重分配 Mask), 因此Attentive Transformer是instance-wise。

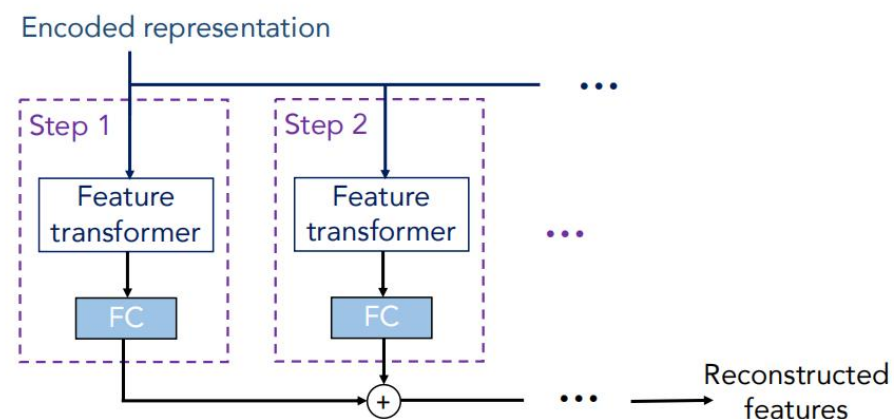
自监督学习

- 自监督学习主要是利用辅助任务 (**pretext**) 从大规模的无监督数据中挖掘自身的监督信息, 通过这种构造的监督信息对网络进行训练, 从而可以学习到对下游任务有价值的表征。
- **TabNet** 对小样本表格数据, 通过随机 **mask** 表格数据中的信息, 进行自监督预训练 (**TabNet encoder**->**TabNet decoder**, 得到表格向量, 再在下游任务中恢复表格数据, 并对其进行预测)

自监督学习, encoder&decoder



(a) TabNet encoder architecture



(b) TabNet decoder architecture

自监督学习，例子

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

TabNet encoder

TabNet decoder

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

将训练好的encoder
交给下游任务使用

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

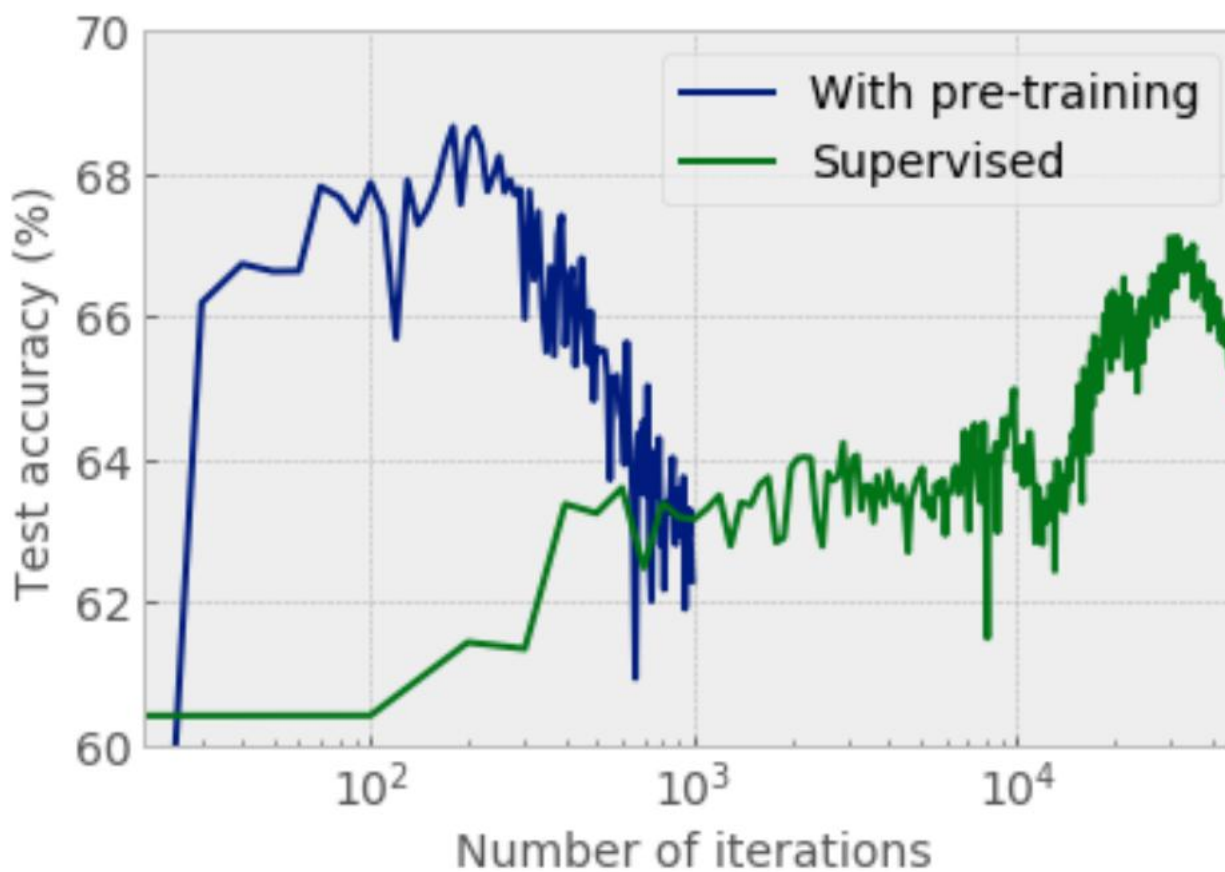
TabNet encoder

Decision making

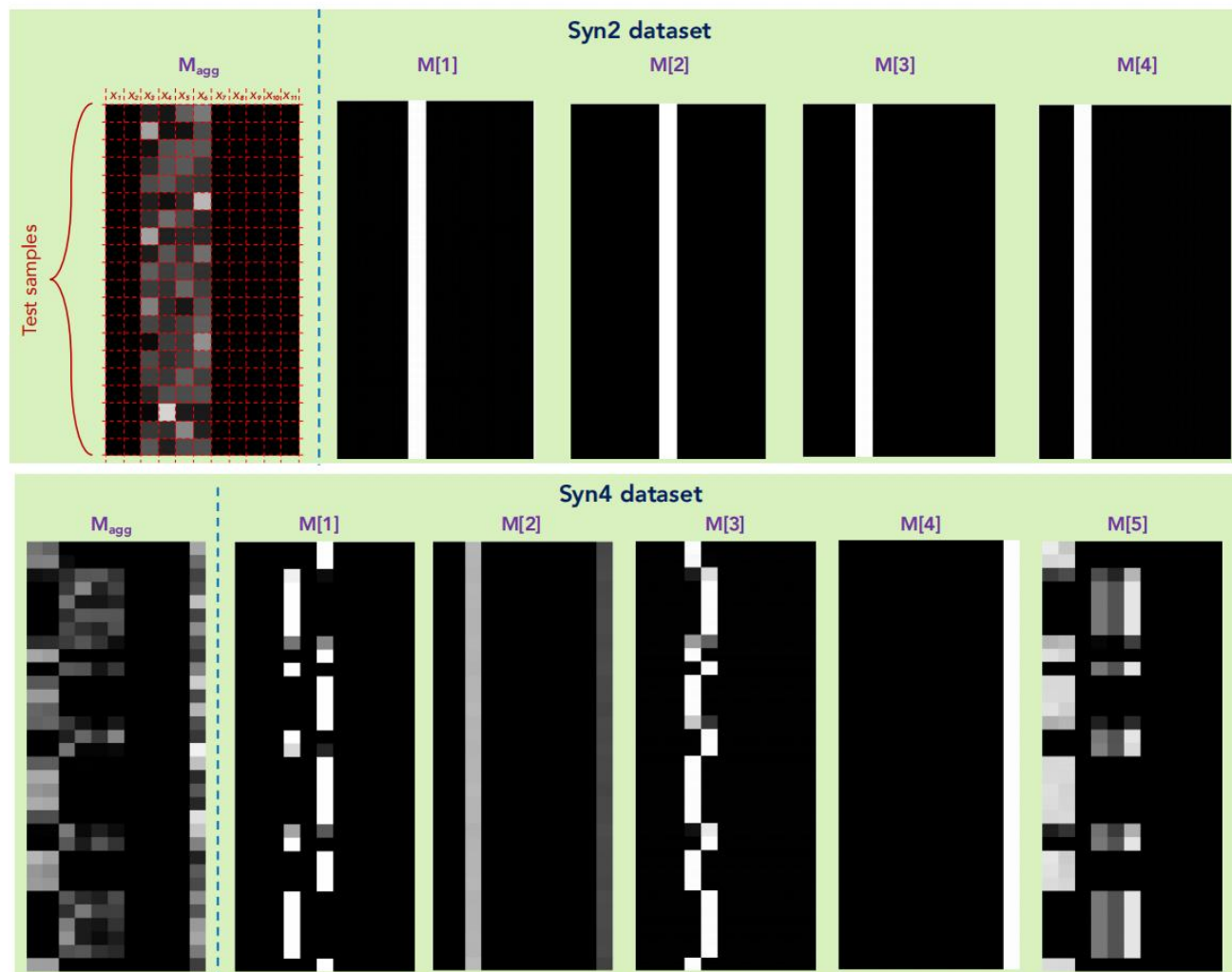
Income > \$50k
True
False
True
False
True
True
False

自监督学习

预训练样本量与准确度的关系



可视化：特征重要性&基于样本的特征选择



可解释性如何实现的呢？

- 答：特征选择+特征计算
- 可输出特征重要度、shap（不确定官方是否提供支持）

总结

- 决策流形的实现：
 - 特征计算：Feature transformer融合了两层：特征共享层和特征独立层
 - 特征选择：Attentive transformer是顺序多步进行的，每个step会根据上一个step的结果来得到当前step选择的特征，并尽量使特征选择结果稀疏且不重复
- 可解释性的实现：
 - TabNet是样本级别的可解释性（每步的特征选择中，特征对每个样本的重要程度，通过可视化显示）
 - 特征重要性
- 自监督学习：
 - encoder + decoder结构
 - 步骤：人为mask->训练encoder，作为下游任务的encoder->下游任务通过已训练encoder的参数作出下游任务决策