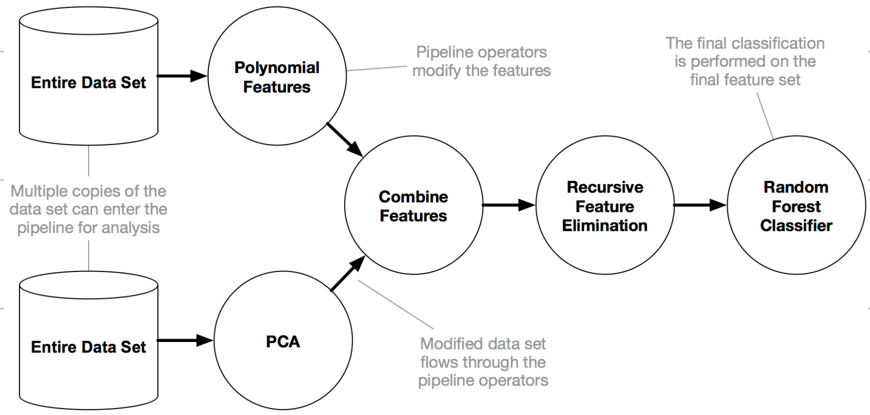
*tpot



tpot

1. 多项式特征筛选（线性回归等）
2. PCA 方差选取
3. 特征合并 ← GP方法迭代，交互熵

（特征内的切分点选取）

4. 使用random froest对结果分类

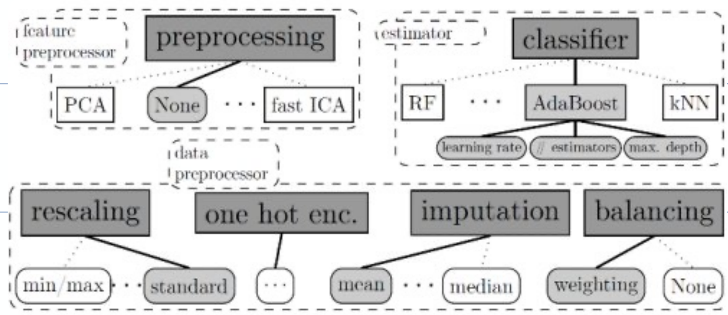
*auto_ml

A tree_based models

N份切分的数据集合⇒特征重要性

B 可前置 Dcn 等模型作特征交叉

* auto_sklearn

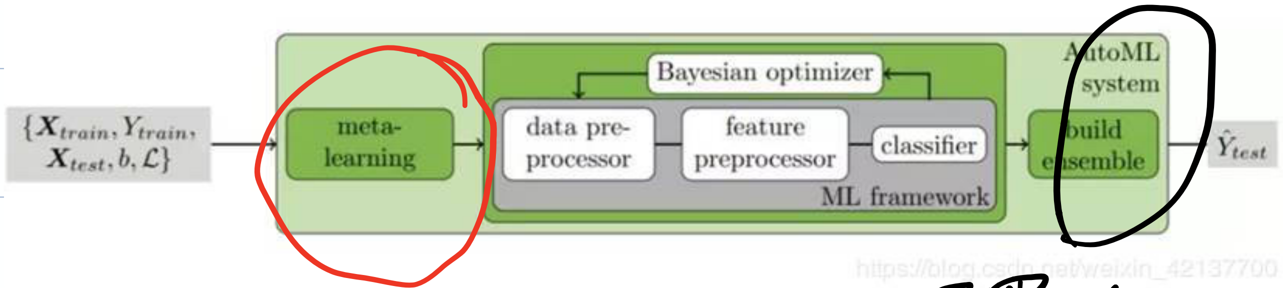


part A：特征预处理（标准化，归一化，平衡，one-hot，缩放 ……）

part B：特征筛选，组合（PCA，独立成分分析）

part C：分类器的效果对比，超参的效果对比

相对来说更暴力，更耗时。



前置元学习

尾置自动化集成功能

元特征：类别数量、特征数量、数据统计信息、目标的熵、数据偏度等，

为整个迭代选取一个接近收敛的初始值。

Name	Formula	Rationale	Variants
Nr instances	n	Speed, Scalability [99]	p/n, $log(n)$, log(n/p)
Nr features	p	Curse of dimensionality [99]	$log(p)$, % categorical
Nr classes	c	Complexity, imbalance [99]	ratio min/maj class
Nr missing values	m	Imputation effects [70]	% missing
Nr outliers	o	Data noisiness [140]	o/n
Skewness	$\frac{E(X-\mu_X)^3}{\sigma_X^3}$	Feature normality [99]	min,max,μ,σ,q_1,q_3
Kurtosis	$\frac{E(X-\mu_X)^4}{\sigma_X^4}$	Feature normality [99]	min,max,μ,σ,q_1,q_3
Correlation	$\rho_{X_1X_2}$	Feature interdependence [99]	min,max,μ,σ,ρ_{XY} [157]
Covariance	$cov_{X_1X_2}$	Feature interdependence [99]	min,max,μ,σ,cov_{XY}
Concentration	$\tau_{X_1X_2}$	Feature interdependence [72]	min,max,μ,σ,τ_{XY}
Sparsity	sparsity(X)	Degree of discreteness [142]	min,max,μ,σ
Gravity	gravity(X)	Inter-class dispersion [5]	
ANOVA p-value	$p_{val_{X_1X_2}}$	Feature redundancy [70]	$p_{val_{XY}}$ [157]
Coeff. of variation	$\frac{\sigma_Y}{\mu_Y}$	Variation in target [157]	
PCA ρ_{λ_1}	$\sqrt{\frac{\lambda_1}{1+\lambda_1}}$	Variance in first PC [99]	$\frac{\lambda_1}{\sum_i \lambda_i}$ [99]
PCA skewness		Skewness of first PC [48]	PCA kurtosis [48]
PCA 95%	$\frac{dim_{95\%var}}{p}$	Intrinsic dimensionality [9]	
Class probability	$P(\mathtt{C})$	Class distribution [99]	min,max,μ,σ
Class entropy	$H(\mathtt{C})$	Class imbalance [99]	
Norm. entropy	$\frac{H(\mathtt{X})}{log_2 n}$	Feature informativeness [26]	min,max,μ,σ
Mutual inform.	$MI(\mathtt{C},\mathtt{X})$	Feature importance [99]	min,max,μ,σ
Uncertainty coeff.	$\frac{MI(\mathtt{C},\mathtt{X})}{H(\mathtt{C})}$	Feature importance [3]	min,max,μ,σ
Equiv. nr. feats	$\frac{H(C)}{MI(C,X)}$	Intrinsic dimensionality [99]	
Noise-signal ratio	$\frac{H(X)-MI(C,X)}{MI(C,X)}$	Noisiness of data [99]	
Fisher's discrimin.	$\frac{(\mu_{c1}-\mu_{c2})^2}{\sigma_{c1}^2-\sigma_{c2}^2}$	Separability classes c_1, c_2 [64]	See [64]
Volume of overlap		Class distribution overlap [64]	See [64]
Concept variation		Task complexity [179]	See [178, 179]
Data consistency		Data quality [76]	See [76]
Nr nodes, leaves	$\|\eta\|, \|\psi\|$	Concept complexity [113]	Tree depth
Branch length		Concept complexity [113]	min,max,μ,σ
Nodes per feature	$\|\eta_X\|$	Feature importance [113]	min,max,μ,σ
Leaves per class	$\frac{\|\psi_c\|}{\|\psi\|}$	Class complexity [49]	min,max,μ,σ
Leaves agreement	$\frac{n_{\psi_i}}{n}$	Class separability [16]	min,max,μ,σ
Information gain		Feature importance [16]	min,max,μ,σ, gini
Landmarker(1NN)	$P(\theta_{1NN}, t_j)$	Data sparsity [115]	Elite 1NN [115]
Landmarker(Tree)	$P(\theta_{Tree}, t_j)$	Data separability [115]	Stump,RandomTree
Landmarker(Lin)	$P(\theta_{Lin}, t_j)$	Linear separability [115]	Lin.Discriminant
Landmarker(NB)	$P(\theta_{NB}, t_j)$	Feature independence [115]	More models [14, 88]
Relative LM	$P_{a,j} - P_{b,j}$	Probing performance [53]	
Subsample LM	$P(\theta_i, t_j, s_t)$	Probing performance [159]	

元学习常见字段