RF Classification and Implementation

Recap

- Discussed XMM variable source classification paper
- Importance of model-fit parameters ,
  - Removing it did not have much effect on accuracy
- NS-BH classification result

Contents

- New improved imputation method
  - Accuracy improved
- Single stage classifier
- Classifier performance
  - Probability quality
  - Accuracy
- Feature Importance method
  - Pair-wise classifier
- Application on real test data

New improved imputation method

Stekhoven, Daniel J., and Peter Bühlmann. "MissForest—non-parametric missing value imputation for mixed-type data." Bioinformatics 28.1 (2011): 112-118

	f1	f2	f3	f4	...	fm
x1				--		
x2		--				
x3		--			--	
x4						--
...	--			--		
xn	--					

New improved imputation method

- Least sparse feature – cadidate column
- Non-candidate > fill with column mean

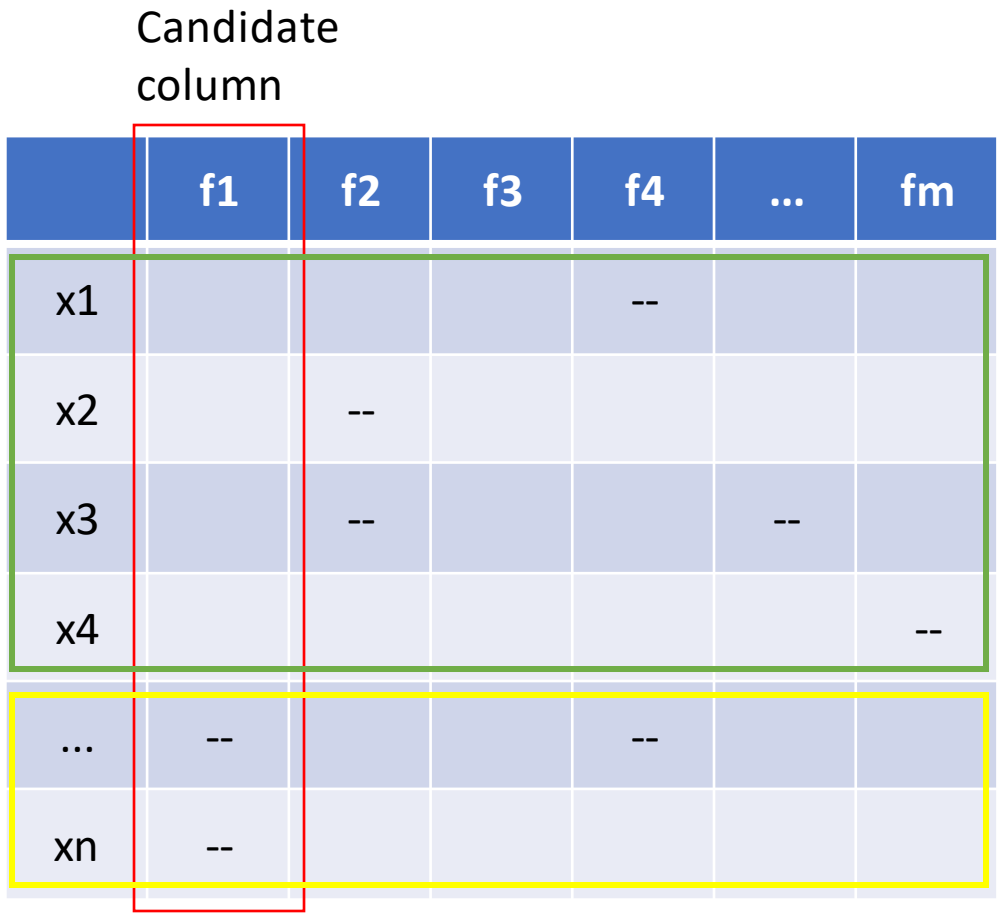
Candidate column

	f1	f2	f3	f4	...	fm
x1				--		
x2		--				
x3		--			--	
x4						--
...	--			--		
xn	--					

Stekhoven, Daniel J., and Peter Bühlmann. "MissForest—non-parametric missing value imputation for mixed-type data." Bioinformatics 28.1 (2011): 112-118

New improved imputation method

- Least sparse feature – cadidate column
- Non-candidate > fill with column mean
- Fit a Regressor
  - Rows with f1 available > training data
  - Impute missing values from regressor
- Move on to next least sparse column.



Training Data

	Num Sources	Num obs
CV	59	508
NS	43	292
BH	24	149
PULSAR	111	288

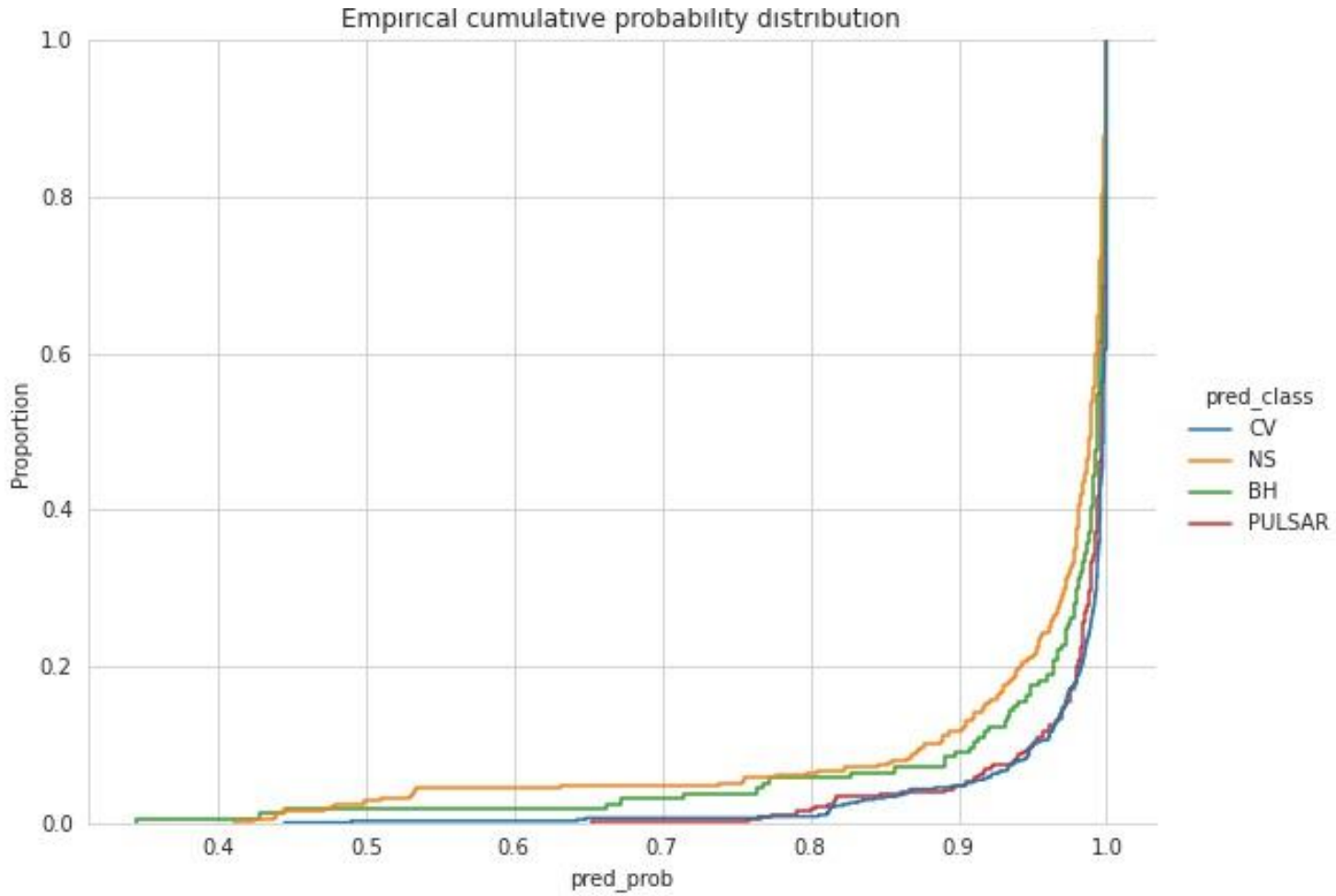
Validation Data

	Num Sources	Num obs
CV	7	8
NS	5	10
BH	4	11
PULSAR	7	31

Accuracy

	Training result	Test result	Validation result
Total samples	1111	124	60
Ambiguous classification	1	7	13
Wrong classification	0	0	0

Probability quality



Application on Sources

CV			NS			BH			Pulsar		
Source Name	**obs**	**clf**	**Source name**	**obs**	**clf**	**Source name**	**obs**	**clf**	**Source name**	**obs**	**clf**
CHI J172550-3533 2	2	CV	3A 2129+470	4	X	CXOU J100518.5-07413	4	BH	PSR J0358+5413 9	9	PL
1WGA J1617.0-2258	2	CV	4U1745-203	2	NS	GINGA 1354-645	3	X	PSR J0437-4715 7	7	PL
IGR J15529-5029	1	CV	EXO 0748-676	2	X	GX 339-4	2	BH	PSR J0418+5732 4	4	PL
CHI J162011-5002	1	X	BW ANT	1	NS	GRO J1655-40	2	X	PSR J0357+3205 4	4	PL
2XMM J231519.0-591029	1	CV	CEN X-4	1	NS				PSR J0205+6449 4	4	PL
XSS J12270-4859	1	X							PSR J0023+0923 2	2	PL
									PSR J0007+7303 1	1	PL

Application on real test data

- Application on 47 TUC GC sources
  - Total number of sources (after filtering)– 265
  - Number of observations – 409
  - Given classification :
    - AB 70
    - CV 60
    - MSP 48
    - AGB 28
    - QLX 10
    - Unknown - 193

- CV
  - cataclysmic variable
- QLX
  - quiescent low-mass X-ray binary containing where accretion onto a neutron star is stopped or greatly reduced
- AB
  - an X-ray active binary consisting of normal stars in a tidally locked short-period binary, where the fast rotation rate drives increased coronal activity
- AGB
  - an AB from Albrow et al. (2001, CDS Cat. <J/ApJ/559/1060>) identified with an X-ray source in this work.
- MSP
  - millisecond radio pulsar

Next Task

- Working on application to 47 TUC
- Try to apply on other GC sources
- Improve network generalisation
- Maybe we can try to source data from other means also – optical / NIR ..

Thank you