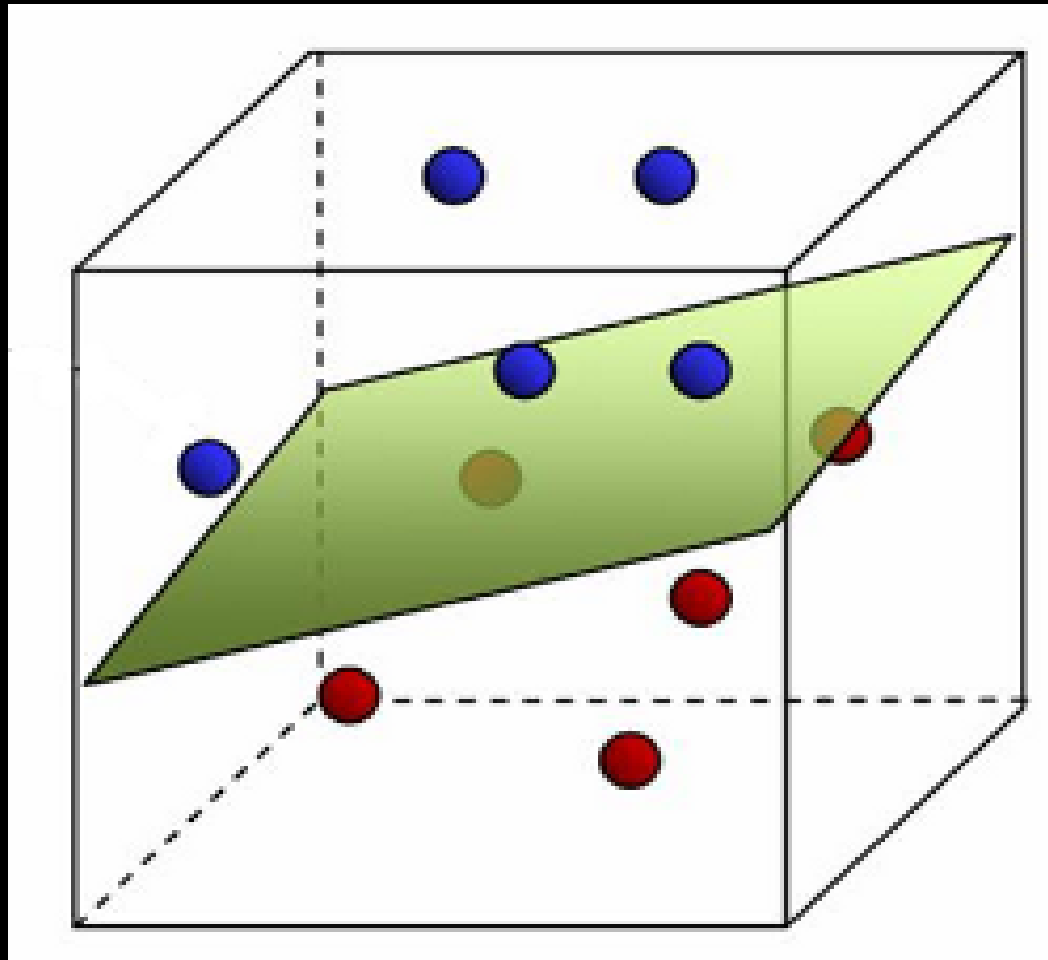


Learning from Partial Information (Two case studies)

Ninan Sajeeth Philip
nspp@iucaa.ernet.in

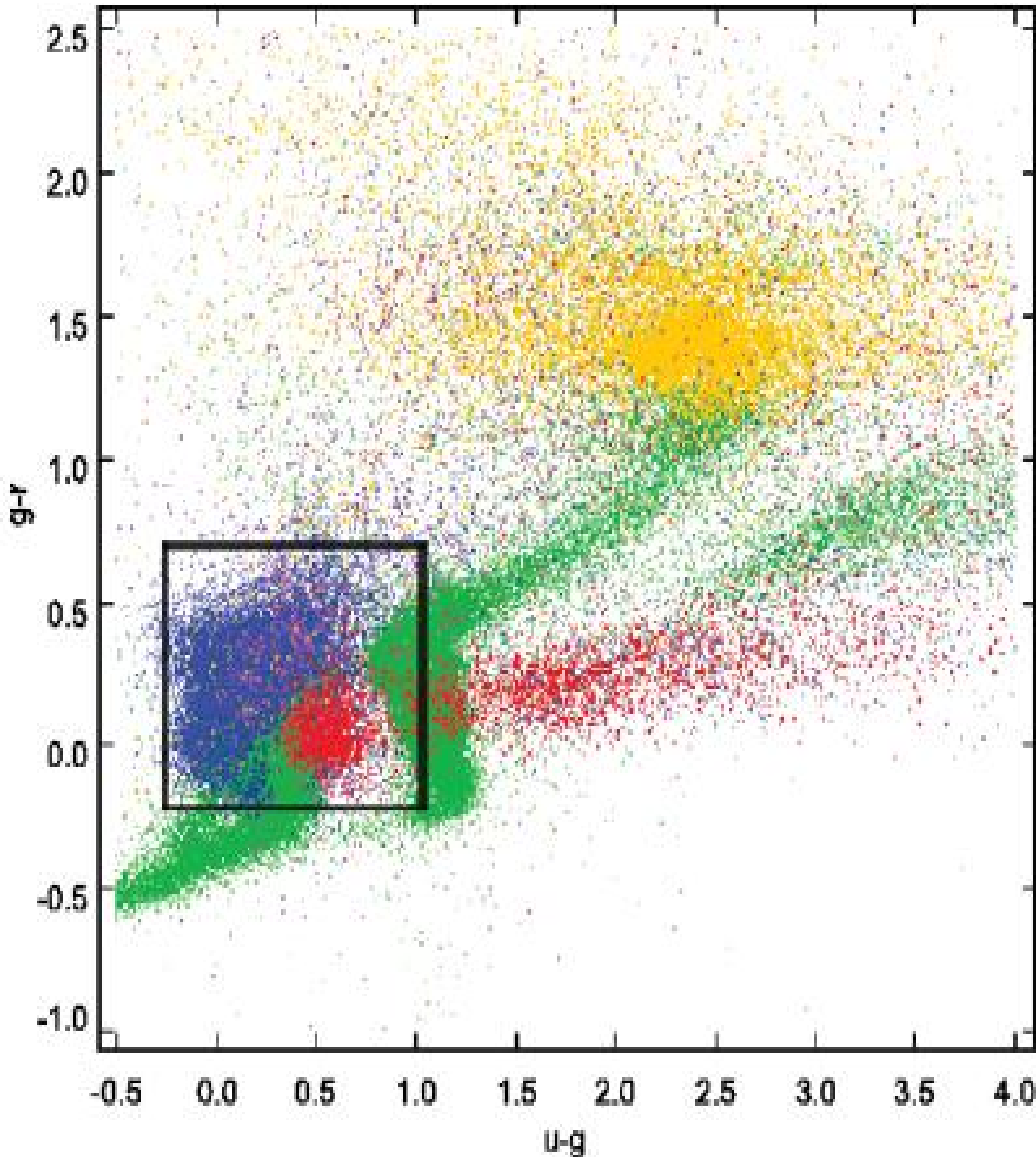
Automated Learning Methods

- All methods assume that the different types (classes) are separable in the feature space.



Light profile of galaxies are different from that of stars

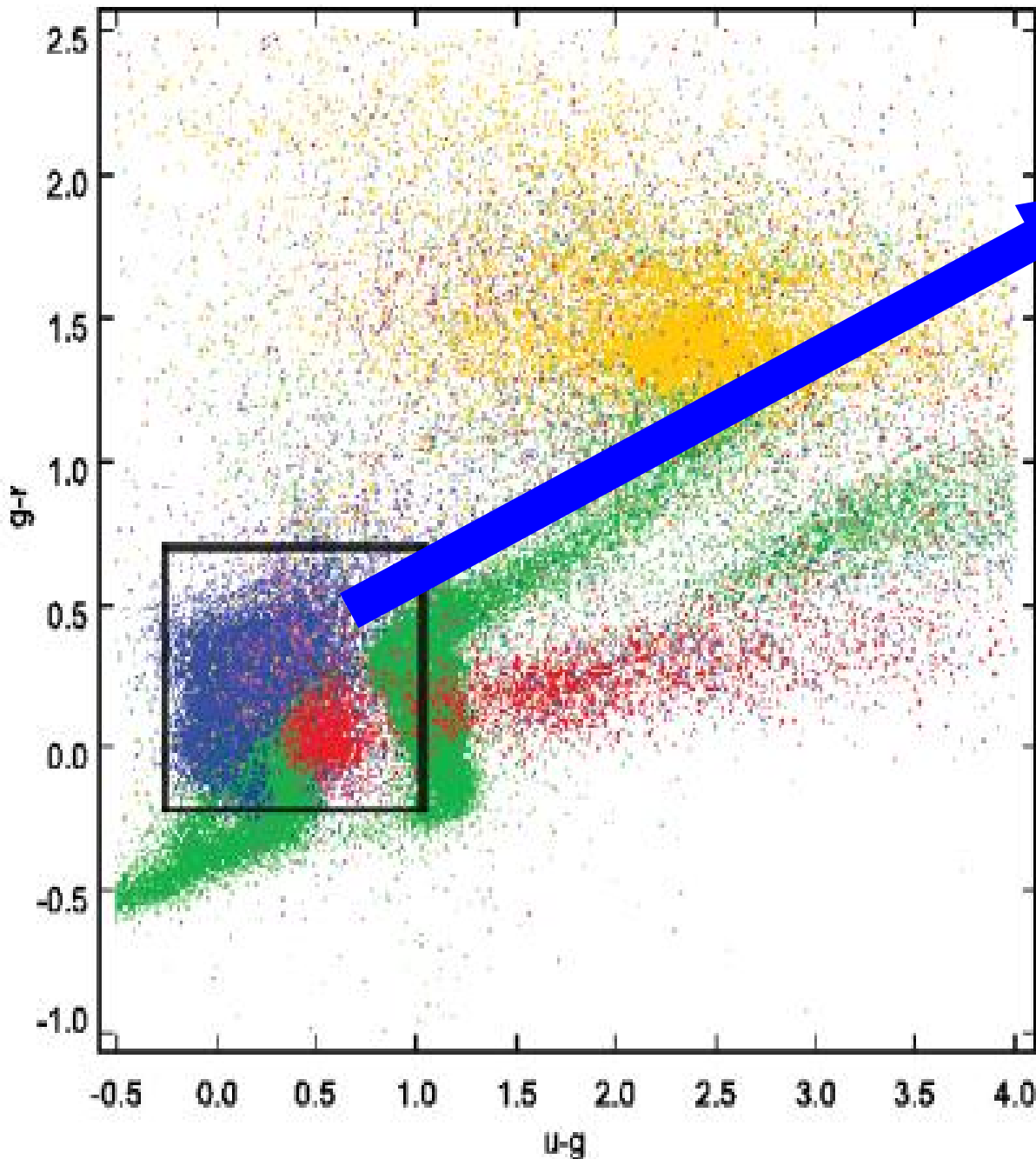
Homogeneous Features: Classification Problem



SDSS
colour-colour plot

Composed of about a million points showing clustering of Quasars (blue and red), main sequence stars (green), late type stars (yellow) and unresolved galaxies (pink) in a colour -colour plot of SDSS colours.

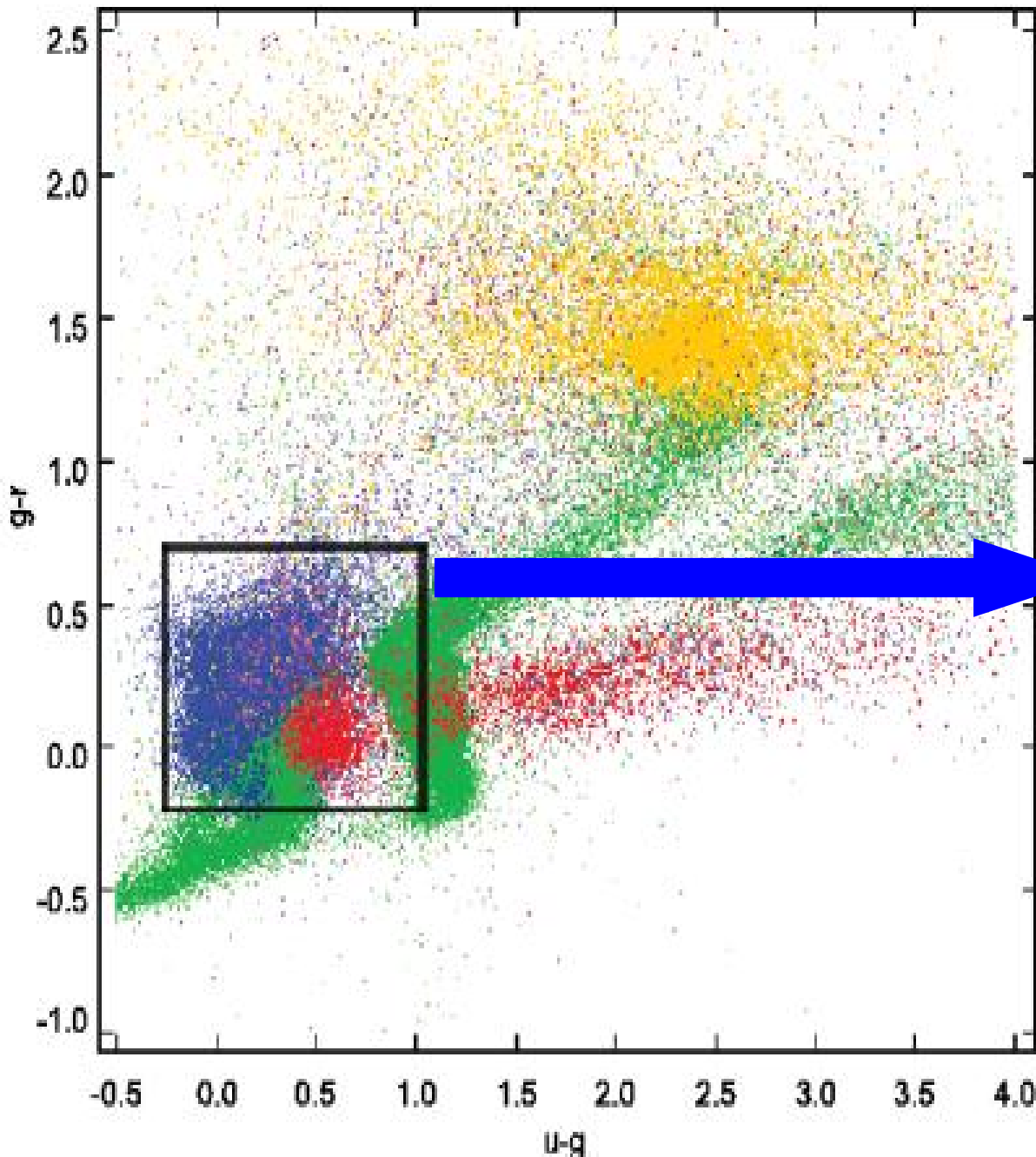
Classification



Blue are low redshift Quasars and our goal is to identify them and verify whether the actual number count match with the estimated values.

Composed of about a million points showing clustering of Quasars (blue and red), main sequence stars (green), late type stars (yellow) and unresolved galaxies (pink) in a colour -colour plot of SDSS colours.

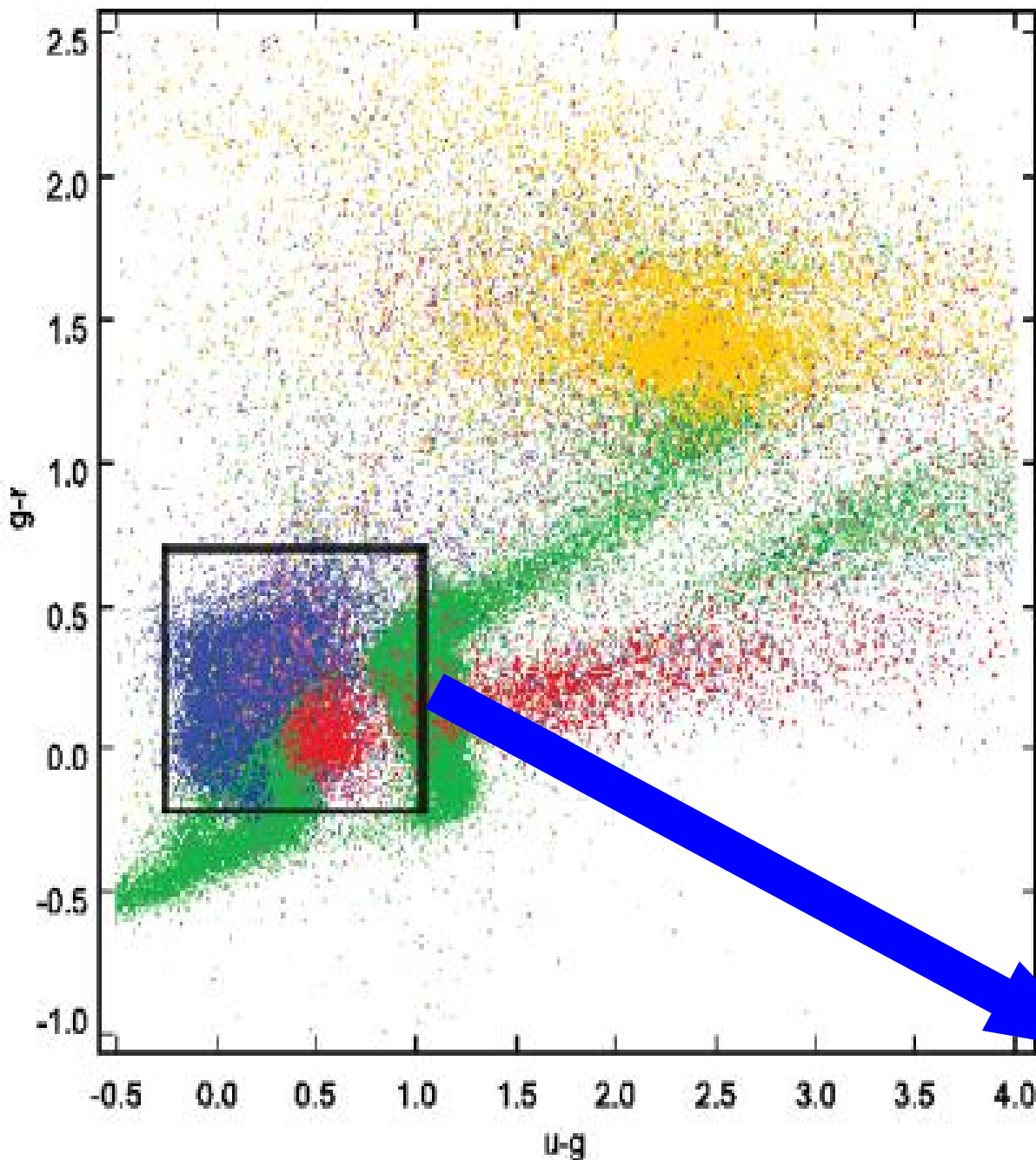
Learn from Examples



Blue are low redshift Quasars and our goal is to identify them and verify whether the actual number count match with the estimated values.

The region in the box has about 150,000 confirmed observations and about 6 million unconfirmed cases.

Map features to Class



Blue are low redshift Quasars and our goal is to identify them and verify whether the actual number count match with the estimated values.

The region in the box has about 150,000 confirmed observations and about 6 million unconfirmed cases.

All objects have known colours (partial information) but the confirmatory spectra and hence class is unknown.

Bayesian Model

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Feature Space

- SDSS provides 5 magnitudes for each object in bands u, g, r, i and z that can be used to construct a ten dimensional colour space.
- A subset of the 150,000 objects with confirmed spectroscopic classification can be used to estimate the likelihood.
- The classifier can be tested on remaining data to verify the accuracy of the model.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Feature Space

- SDSS provides 5 magnitudes for each object in bands u, g, r, i and z that can be used to construct a ten dimensional colour space.
- A subset of the 150,000 objects with confirmed spectroscopic classification can be used to estimate the likelihood. The distribution is not smooth
- The classifier can be tested on remaining data to verify the accuracy of the model.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Feature Space

- SDSS provides 5 magnitudes for each object in bands u, g, r, i and z that can be used to construct a ten dimensional colour space.
- A subset of the 150,000 objects with confirmed spectroscopic classification can be used to estimate the likelihood.
~~The distribution is not smooth~~
The colour space need to be binned to approximate the distribution.
- The classifier can be tested on remaining data to verify the accuracy of the model.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Feature Space

- SDSS provides 5 magnitudes for each object in bands u, g, r, i and z that can be used to construct a ten dimensional colour space.
- A subset of the 150,000 objects with confirmed spectroscopic classification can be used to estimate the likelihood.

The distribution is not smooth

The colour space need to be binned to approximate the distribution. Computing conditional likelihood of the binned high dimensional feature space is nearly impossible.

- The classifier can be tested on remaining data to verify the accuracy of the model.

Two issues with Bayesian Formalism

- How would you guess the True value of the **Prior** for each bin?
- **Conditional dependency** of the input feature space – likelihood is conditionally dependent on feature vectors - Naive Bayesian models that ignore conditional dependence fail on even simple XOR problems.

Bayesian Prior

- **Ensemble methods**: Multiple models, same data : many weak learners combined to form a strong learning model
- Bagging: each model in ensemble vote for the probable candidate
- Boosting: Emphasise the failing models with weights
- Bayesian Model Averaging (BMA): Sampling Hypothesis from Hypothesis Space
- Bayesian Model Combination (BMC): Seek combination of models closest to a distribution.

Bayesian Methods

- **Ensemble methods**: Multiple models, same data
- **Bagging**: each model in ensemble **vote** for the probable candidate
- **Boosting**: Emphasise the failing models with weights
- **Bayesian Model Averaging (BMA)**: Sampling Hypothesis from Hypothesis Space
- **Bayesian Model Combination (BMC)**: Seek combination of models closest to a distribution.

Bayesian Methods

- **Ensemble methods**: Multiple models, same data
- Bagging: each model in ensemble vote for the probable candidate
- **Boosting**: Emphasise the failing models with **weights**
- Bayesian Model Averaging (BMA): Sampling Hypothesis from Hypothesis Space
- Bayesian Model Combination (BMC): Seek combination of models closest to a distribution.

Bayesian Methods

- **Ensemble methods**: Multiple models, same data
- Bagging: each model in ensemble vote for the probable candidate
- Boosting: Emphasise the failing models with weights
- **Bayesian Model Averaging (BMA)**: Sampling Hypothesis from Hypothesis Space
- Bayesian Model Combination (BMC): Seek combination of models closest to a distribution.

Bayesian Methods

- **Ensemble methods**: Multiple models, same data
- Bagging: each model in ensemble vote for the probable candidate
- Boosting: Emphasise the failing models with weights
- Bayesian Model Averaging (BMA): Sampling Hypothesis from Hypothesis Space
- **Bayesian Model Combination (BMC)**: Seek a combination of models closest to a distribution.

Our Solution

- Estimating both Prior and Likelihood from data.
- **Boosting:** Emphasise the failing models with weights

Replace Prior by weights within the same model.

Likelihood estimation of Binned Space

- There may not be sufficient samples in each bin to estimate likelihood when conditional dependence constraints are imposed on them.
- We adopted an imposed conditional independence formula that approximate the likelihood for a conditionally dependent event as the product of the likelihood for pairs of input features.

Likelihood estimation of Binned Space

- There may not be sufficient samples to estimate likelihood when constraints on conditional dependence is imposed on them.
- We adopted an **imposed conditional independence** formula that approximate the likelihood for a conditionally dependent event.

Imposed Conditional Independence

The likelihood for a conditionally dependent event A is approximated as

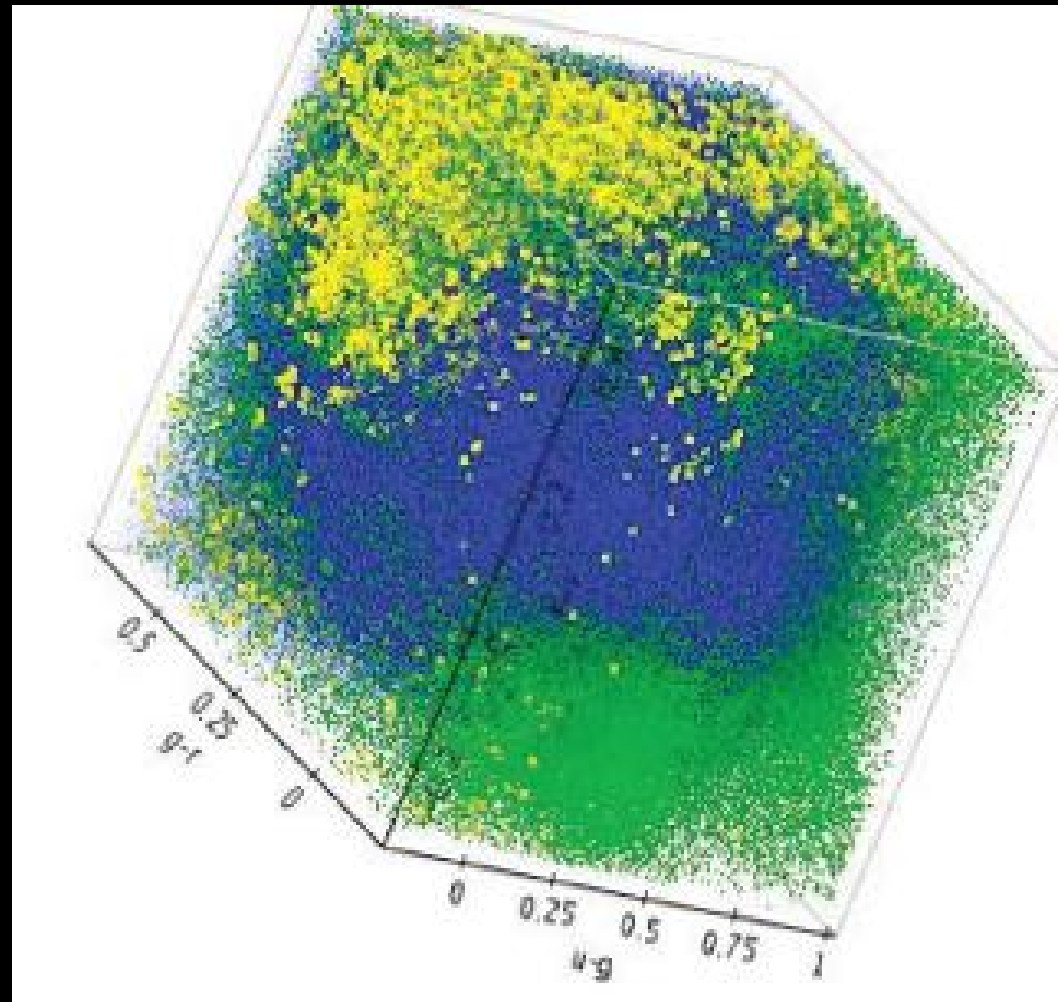
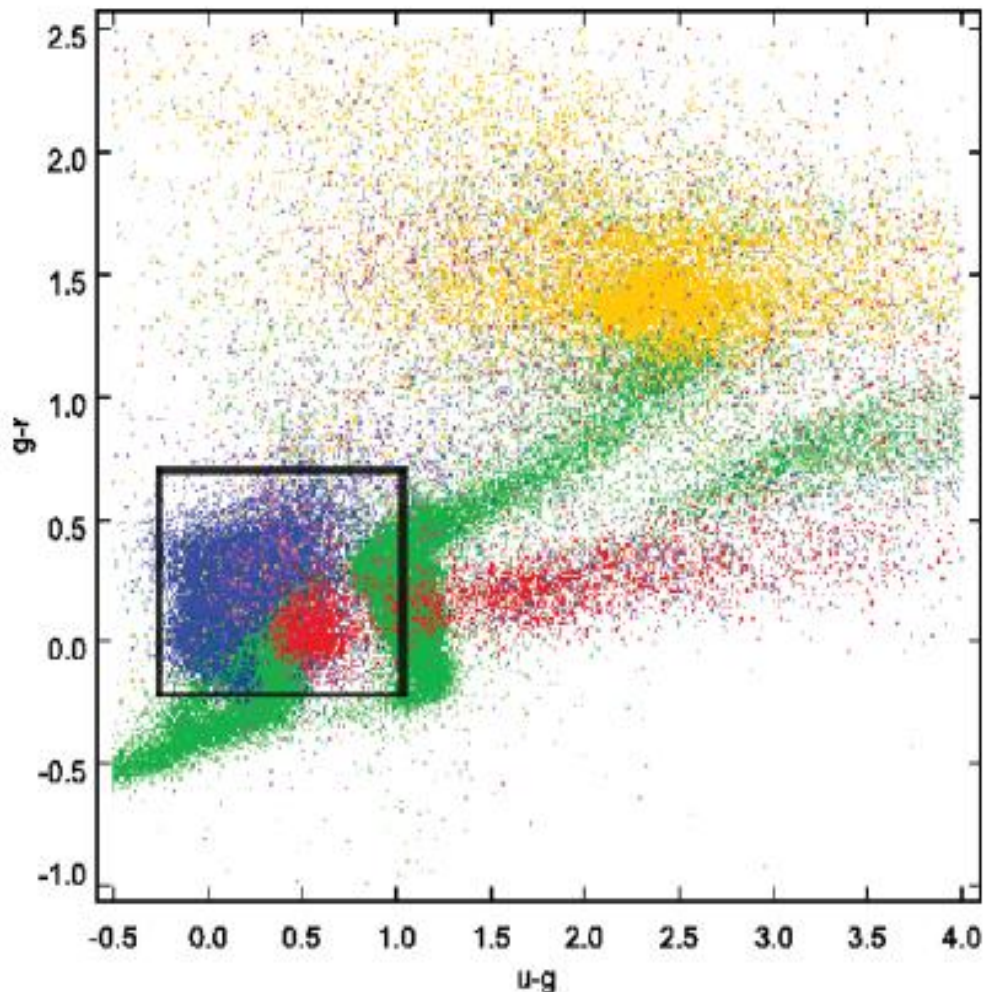
- $L(A|b,c,d,e,f) \sim$
 $M * L(A|b,c) * L(A|b,d) * L(A|b,e) * L(A|b,f) * L(A|c,d) * L(A|c,e) * L(A|c,f) * L(A|d,e) * L(A|d,f) * L(A|e,f)$
- Works better than Naive Bayes – no issue with XOR gate

Imposed Conditional Independence

The likelihood for a conditionally dependent event A can be approximated as the product of the likelihood of its paired inputs.

- $L(A|b,c,d,e,f) \sim$
 $M * L(A|b,c) * L(A|b,d) * L(A|b,e) * L(A|b,f) * L(A|c,d) * L(A|c,e) * L(A|c,f) * L(A|d,e) * L(A|d,f) * L(A|e,f)$
- Works better than Naive Bayes – no issue with XOR gate

Classification of the 6 million Objects



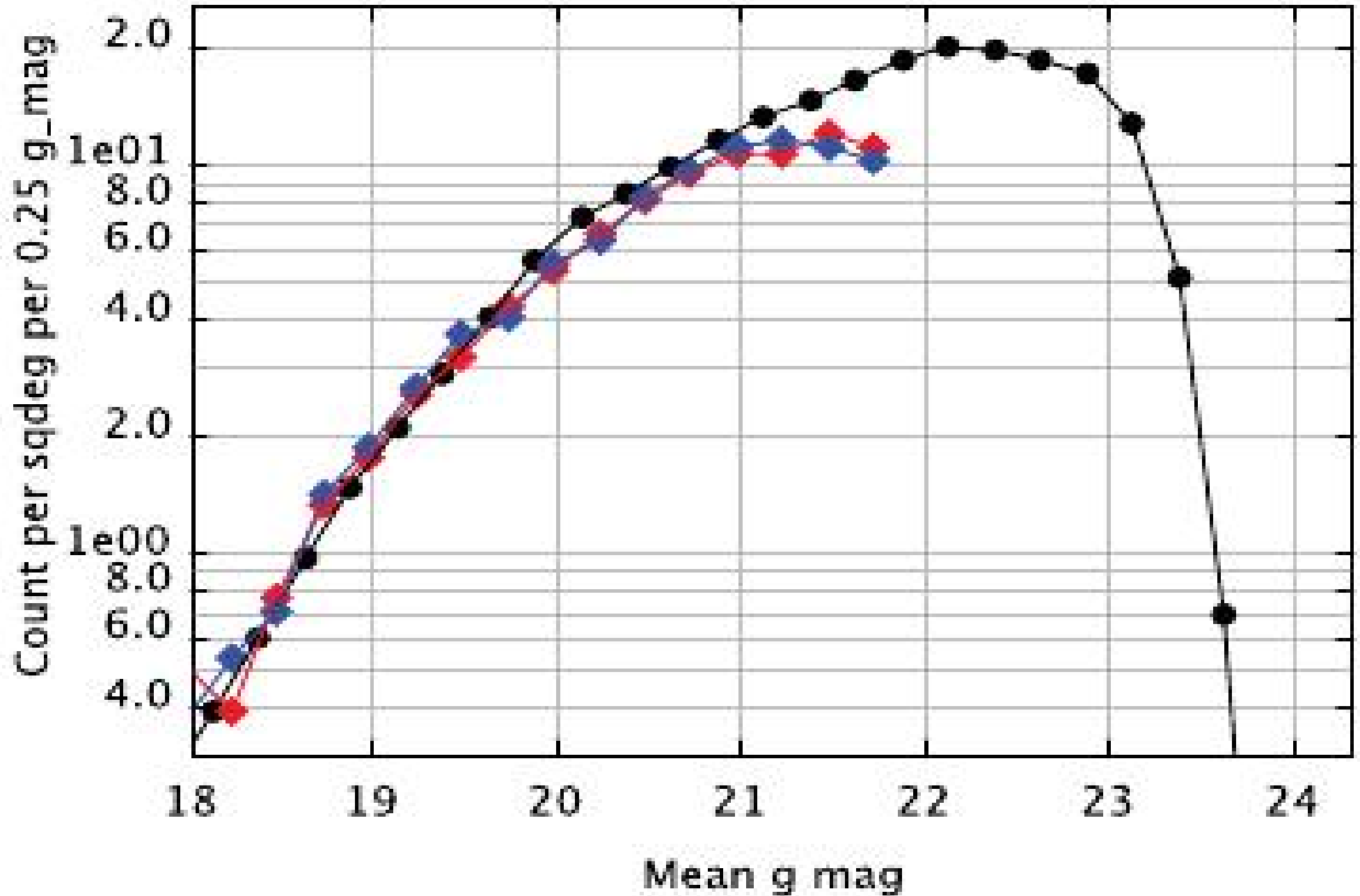
Blue are Quasars, Yellow are unresolved Galaxies and Green are main sequence Stars

Verification using Heterogeneous Surveys

Table 5. Summary of the matching of our catalogue predictions with some existing catalogues.

Cat. code	DBNN predictions			Failures as per catalogue			Accuracy (per cent)	<i>i</i> mag range	Ref
	Quasar	Galaxy	Star	Quasar	Galaxy	Star			
2DF	5976	238	1535	122	0	52	98	17.0–22.0	1
XBH	212	0	0	0	0	0	100	15.8–20.5	2
ASFS	1088	12	31	0	12	31	96	14.5–22.1	3
BATCS	21	0	0	3	0	0	86	18.1–20.5	4
CGRBS	265	1	0	0	1	0	100	14.7–21.5	5
DLyaQ	21	0	1	0	0	1	95	16.5–19.4	6
F2QZ	186	1	3	0	1	3	98	16.6–21.0	7
KFQS	144	2	13	3	1	7	94	16.8–20.6	8
LQAC	61 504	17	267	0	17	267	100	14.7–22.3	9
LQRF	60 280	14	219	0	14	219	100	14.7–21.7	10
BZC	249	4	2	0	4	2	98	15.0–21.0	11
PCS	53	0	2	0	0	2	96	15.1–18.5	12
ROSA	1134	0	1	0	0	1	100	15.5–20.5	13
SQ13	65 223	55	395	0	55	395	99	14.7–22.8	14
SQR13	7	0	21	7	0	0	75	16.3–20.3	14
DR7Q	79 140	17	341	0	17	341	100	14.9–21.8	15
SSSC	82	2	1171	82	2	0	93	14.9–21.5	16
SSA13	5	0	1	0	0	0	83	17.8–20.8	17
XMMSS	37	0	5	1	0	2	93	14.9–20.7	18
SDSS/XMM	580	0	0	0	0	0	100	15.2–20.5	19
RASS/2MASS	6	0	0	0	0	0	100	15.5–18.4	20
CAIXA	16	0	0	0	0	0	100	15.1–17.8	21
WDMB	20	0	106	20	0	0	84	15.3–20.5	22
PMS	639	6	19 596	639	6	0	97	14.8–20.2	23
GLQ	2	0	0	0	0	0	100	18.8–19.1	24

Comparison with expected number counts



Further Information



A photometric catalogue of quasars and other point sources in the Sloan Digital Sky Survey

Sheelu Abraham,^{1★} Ninan Sajeeth Philip,^{1★} Ajit Kembhavi,^{2★} Yogesh G. Wadadekar^{3★}
and Rita Sinha^{4★†}

¹*St Thomas College, Kozhencheri 689641, India*

²*Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411007, India*

³*National Centre for Radio Astrophysics, TIFR, Post Bag 3, Ganeshkhind, Pune 411007, India*

⁴*Elviraland 194, 2591 GM The Hague, the Netherlands*

The Predicted Catalogue

Monthly Notices

of the

ROYAL ASTRONOMICAL SOCIETY



Mon. Not. R. Astron. Soc. **419**, 80–94 (2012)

doi:10.1111/j.1365-2966.2011.19674.x

A photometric catalogue of quasars and other point sources in the Sloan Digital Sky Survey

Sheelu Abraham,¹★ Ninan Sajeeth Philip,¹★ Ajit Kembhavi,²★ Yogesh G. Wadadekar³★
and Rita Sinha⁴★†

¹*St Thomas College, Kozhencheri 689641, India*

²*Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411007, India*

³*National Centre for Radio Astrophysics, TIFR, Post Bag 3, Ganeshkhind, Pune 411007, India*

⁴*Elviraland 194, 2591 GM The Hague, the Netherlands*

Search Criteria
Save in CDSportal
Keywords Back
J/MNRAS/419/80
Tables Add
J/MNRAS/419/80
..catalog

Portal Simbad VizieR Aladin X-Match Other Help

Catalog Selection Page

Simple Target List Of Targets [Fast Xmatch with large catalogs or Simbad](#)

Target Name (resolved by [Sesame](#)) or Position: Target dimension:
Clear [J2000] 2 arcmin
 Radius Box size

Photometric Classification Catalogue of SDSS DR7 (Abraham+, 2012) [Similar Catalogs](#) [ReadMe+ftp](#)

J/MNRAS/419/80 [Post annotation](#)

1.J/MNRAS/419/80/catalog Photometric catalogue based on SDSS DR7 (a sample is published as Table 4) (6038247 rows)

A more complex situation

- What if **all input features are not known?**

Straightforward solution : Compute the inverse probability for the missing feature just as you handle missing values.

Not so easy situation: What if we do not have a training data with all features for computing inverse probability?

Heterogeneous Input Features

- What if all input features are not known?

Straightforward answer : Compute the inverse probability for the missing feature just as you handle missing values.

Not so easy situation: What if we do not have a training data with all features for computing inverse probability?

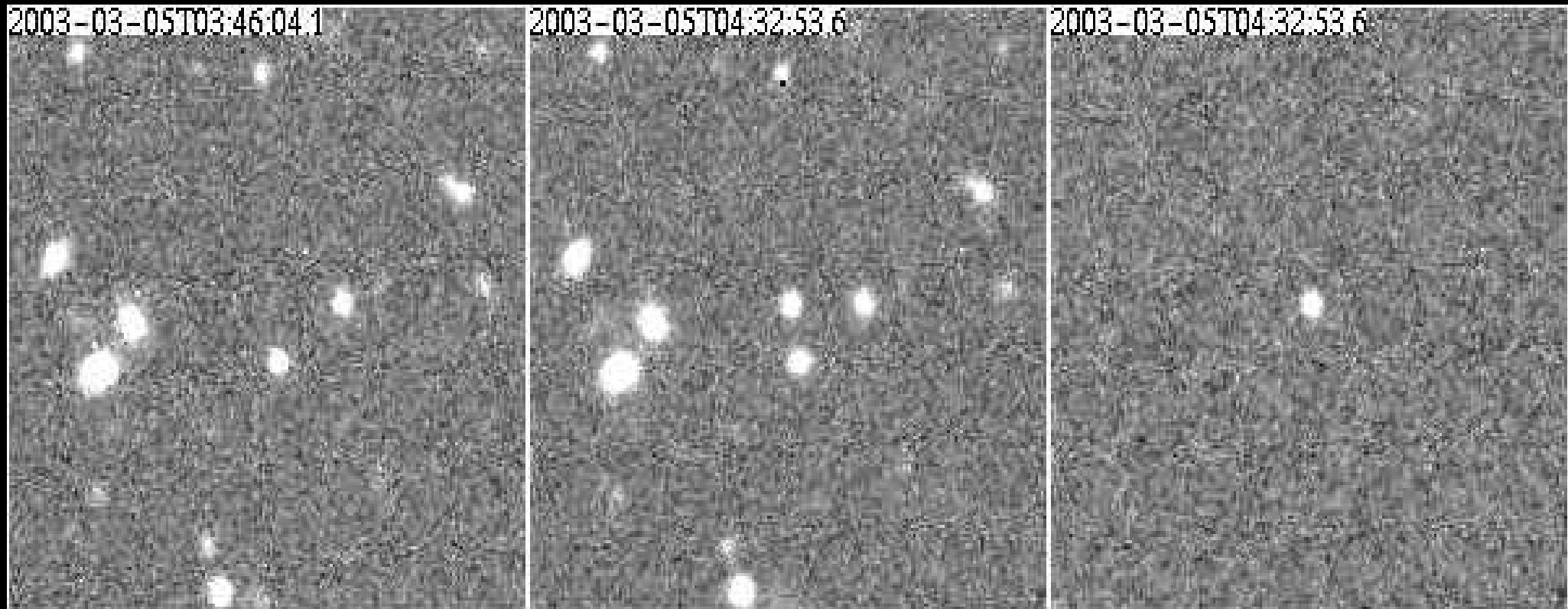
A more complex situation

- What if all input features are not known?

Straightforward answer : Compute the inverse probability for the missing feature just as you handle missing values.

Not so easy situation: If we do not have a training data with all features for computing inverse probability.

A Challenging Problem



CRTS

*Catalina Real-Time
Transient Survey*

- All Transients
 - CSS(*large*)
 - MLS(*large*)
 - SSS
- Supernovae
 - SN Hunt
 - CSS
 - MLS
 - SSS
- Blazars
 - CSS
 - MLS
 - SSS
- Bright CVs
 - CSS
 - MLS
 - SSS
- All CVs
 - CSS
 - MLS
 - SSS
- AGN
 - CSS
 - MLS
 - SSS
- Flares and Asts
 - CSS
 - MLS
 - SSS
- SN or CVs
 - CSS
 - MLS
 - SSS
- Others types
 - CSS
 - MLS
 - SSS

A Challenging Problem

Generate alerts on optical transient detections

Minimize false alarms

Customize alarms to user demands

Send the alarms immediately – given minimal or sometime very little information about it.

Example : Nearest distance to a galaxy or star

: Distance to nearest known radio object

: Distance to nearest known x-ray detections

: Magnitudes in archives and in earlier detections

All Transients

- CSS(large)
- MLS(large)
- SSS

Supernovae

- SN Hunt
- CSS
- MLS
- SSS

Blazars

- CSS
- MLS
- SSS

Bright CVs

- CSS
- MLS
- SSS

All CVs

- CSS
- MLS
- SSS

AGN

- CSS
- MLS
- SSS

Flares and Asts

- CSS
- MLS
- SSS

SN or CVs

- CSS
- MLS
- SSS

Others types

- CSS
- MLS
- SSS

A Challenging Problem

Generate alerts on optical transient detections

Minimize false alarms

Customize alarms to user demands

Send the alarms immediately – given minimal or sometime very little information about it.

Example : Nearest distance to a galaxy or star

: Distance to nearest known radio object

: Distance to nearest known x-ray detections

: Magnitudes in archives and in earlier detections

A Challenging Problem

Generate alerts on optical transient detections

Minimize false alarms

Customize alerts to user demands

Send the alarms immediately – given minimal or sometime very little information about it.

Example : Nearest distance to a galaxy or star

: Distance to nearest known radio object

: Distance to nearest known x-ray detections

: Magnitudes in archives and in earlier detections

All Transients

- CSS (large)
- MLS (large)
- SSS

Supernovae

- SN Hunt
- CSS
- MLS
- SSS

Blazars

- CSS
- MLS
- SSS

Bright CVs

- CSS
- MLS
- SSS

All CVs

- CSS
- MLS
- SSS

AGN

- CSS
- MLS
- SSS

Flares and Asts

- CSS
- MLS
- SSS

SN or CVs

- CSS
- MLS
- SSS

Others types

- CSS
- MLS
- SSS

A Challenging Problem

Generate alerts on optical transient detections

Minimize false alarms

Customize alarms to user demands

Send the alerts immediately – given minimal or sometime very little information about it.

Example : Nearest distance to a galaxy or star

: Distance to nearest known radio object

: Distance to nearest known x-ray detections

: Magnitudes in archives and in earlier detections

- All Transients
 - CSS (large)
 - MLS (large)
 - SSS
- Supernovae
 - SN Hunt
 - CSS
 - MLS
 - SSS
- Blazars
 - CSS
 - MLS
 - SSS
- Bright CVs
 - CSS
 - MLS
 - SSS
- All CVs
 - CSS
 - MLS
 - SSS
- AGN
 - CSS
 - MLS
 - SSS
- Flares and Asts
 - CSS
 - MLS
 - SSS
- SN or CVs
 - CSS
 - MLS
 - SSS
- Others types
 - CSS
 - MLS
 - SSS

All Transients

- CSS(large)
- MLS(large)
- SSS

Supernovae

- SN Hunt
- CSS
- MLS
- SSS

Blazars

- CSS
- MLS
- SSS

Bright CVs

- CSS
- MLS
- SSS

All CVs

- CSS
- MLS
- SSS

AGN

- CSS
- MLS
- SSS

Flares and Asts

- CSS
- MLS
- SSS

SN or CVs

- CSS
- MLS
- SSS

Others types

- CSS
- MLS
- SSS

A Challenging Problem

Generate alerts on optical transient detections

Minimize false alarms

Customize alarms to user demands

Send the alerts immediately – given minimal or sometime very little information about it.

Example : Nearest distance to a galaxy or star

: Distance to nearest known radio object

: Distance to nearest known x-ray detections

: Magnitudes in archives and in earlier detections

Missing Values

Example : Nearest distance to a galaxy or star

: Distance to nearest known radio object

: Distance to nearest known x-ray detections

: Magnitudes in archives and in earlier detections

Possible only if the object is within the foot print of a survey

Each survey may use a different unit for their catalogues
– need to be considered separately

Missing Values

Example : Nearest distance to a galaxy or star

: Distance to nearest known radio object

: Distance to nearest known x-ray detections

: Magnitudes in archives and in earlier detections

Possible only if the object is within the foot print of a survey

Each survey may use a different unit for their catalogues
– need to be considered separately

Missing Data Values

```
1 | 0.2593009999999997 0.3348999999999998 0.4155999999999998 0.07559900000000004 0.1562990000000001
   | 0.08070000000000002 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 1.3 -9999 -9999
   | 22.7 -9999 -9999 -9999 23.2 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 3 -9999 -9999 -9999 2
2 | -0.0428999999999995 -0.0269999999999992 -0.0239999999999991 0.01590000000000002 0.01890000000000004
   | 0.003000000000000011 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 17.2 -9999 -9999
   | 29.8 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 11
3 | 0.0260999999999996 0.05920000000000006 0.0274999999999999 0.03310000000000001 0.001400000000000029
   | -0.03170000000000007 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 9.3 -9999 -9999
   | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 11
4 | -0.0411999999999999 0.170099 0.01199899999999994 0.211299 0.05319899999999993 -0.15810000000000001 -9999 -9999
   | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 4.8 -9999 -9999 -9999 -9999 -9999 -9999 -9999
   | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 2
5 | 0.25090000000000001 -0.1667999999999999 0.18240000000000001 -0.4177 -0.06850000000000002 0.3492 -9999 -9999 -9999
   | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 13.3 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999
   | -9999 6.4 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 6.37076376235 14.9457997427 2
6 | 0.3771989999999997 0.2477989999999997 0.1267 -0.1294 -0.2504989999999998 -0.1210989999999997 -9999 -9999 -9999
   | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 0.8 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999
   | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 2
7 | 0.791 -0.08769999999999981 0.20220000000000001 -0.8786999999999998 -0.5887999999999999 0.2898999999999999 -9999
   | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 1.4 -9999 -9999 -9999 -9999 -9999 -9999
   | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 6
```

The training data itself has missing data values.

Note: The accuracy of the actual observation is not beyond one or two decimal places. The double precision is used here only to reduce round off error while rescaling the data during the processing.

Missing Values

```
1 | 0.2593009999999997 0.3348999999999998 0.4155999999999998 0.07559900000000004 0.1562990000000001
  | 0.08070000000000002 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 1.3 -9999 -9999
  | 22.7 -9999 -9999 -9999 23.2 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 3 -9999 -9999 -9999 2
2 | -0.0428999999999995 -0.0269999999999992 -0.0239999999999991 0.01590000000000002 0.0000000000000000
  | 0.003000000000000011 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 2 -9999 -9999
  | 29.8 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 11
3 | 0.0260999999999996 0.05920000000000006 0.0274999999999999 0.03310000000000001 0.00000000000000029
  | -0.03170000000000007 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 9.3 -9999 -9999
  | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 11
4 | -0.0411999999999999 0.170099 0.0119989999999999 0.211299 0.05310000000000003 0.08100000000000001 -9999 -9999
  | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 4.8 -9999 -9999 -9999 -9999
  | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 2
5 | 0.25090000000000001 -0.1667999999999999 0.18210000000000001 -0.41110000000000002 0.06850000000000002 0.3492 -9999 -9999 -9999
  | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 6.4 -9999 -9999 -9999 -9999
  | -9999 6.4 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 6.37076376235 14.9457997427 2
6 | 0.37719899999999997 0.24779999999999997 0.12000000000000001 0.1294 -0.25049899999999998 -0.12109899999999997 -9999 -9999 -9999
  | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 0.8 -9999 -9999 -9999 -9999 -9999 -9999 -9999
  | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 2
7 | 0.791099999999999981 0.20220000000000001 -0.87869999999999998 -0.5887999999999999 0.28989999999999999 -9999
  | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 1.4 -9999 -9999 -9999 -9999 -9999
  | -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 6
```

No way to compute inverse probability - makes it impossible for standard machine learning algorithms to learn and predict the outcome

Our Approach

The likelihood for a conditionally dependent event A can be approximated as the product of the likelihood of its paired inputs.

- $L(A|b,c,d,e,f) \sim$
 $M * L(A|b,c) * L(A|b,d) * L(A|b,e) * L(A|b,f) * L(A|c,d) * L(A|c,e) * L(A|c,f) * L(A|d,e) * L(A|d,f) * L(A|e,f)$

Our Approach

The likelihood for a conditionally dependent event A can be approximated as the product of the likelihood of its paired inputs.

- $L(A|b,c,d,e,f) \sim$

$$M * L(A|b,c) * L(A|b,d) * L(A|b,e) * L(A|b,f) * L(A|c,d) * L(A|c,e) * L(A|c,f) * L(A|d,e) * L(A|d,f) * L(A|e,f)$$

- Estimate approximate Likelihood based on whatever information available and use it for training and testing.

Dynamic Learning

- With lot of missing values in the observations, each input data has partial information about the features associated to an outcome.
- Learn as we go... use Bayesian update rule to update the belief in each input feature and its consequences.

Dynamic Learning

- With lot of missing values in the observations, each input data has partial information about the features associated to an outcome.
- Learn as we go... use **Bayesian update rule** to update the belief in each input feature and its effect on the outcome.

Dynamic Addition of Features

- We want to use all available information about the detections as and when they become available.
- Since likelihood is computed as the product, it is feasible to update it with new evidences as and when they become available.

Dynamic Addition of Features

- We want to use all available information about the detections
- Since likelihood is computed as the product, it is feasible to update it with new evidences as and when they become available.

Dreaming Computers

- We now have many input features but not so many examples to learn from. This can lead to over-fitting the data and Memorising rather than generalising the situation.
- Dreams are synthetic inputs our brain uses to teach us how to react to plausible situations. Can we create dreams for computers?

Dreaming Computers

- We now have many input features but not so many examples to learn from. This can lead to over-fitting the data and Memorising rather than generalising the situation.
- **Hypothesis:** Dreams are synthetic inputs our brain uses to teach us how to react to plausible situations. **Can we create dreams for computers?**

Information from Error Bars

- Error bars tell us that the nature of the object remains same even if the measurement value is perturbed within the range of the error bar – can be used to generate new data

DBNN Annotator



A collaborative project with Ashish Mahabal (Caltech), IUCAA, Pune and the CRTS Team with funding from IUSSTF and ISRO.

CRTS Predictions

	[1]	[2]	[3]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[16]	Total	
[1]	273	3	4	1	1	0	1	3	3	3	0	1	293	1, "Cataclysmic Variable"
[2]	4	402	3	0	4	1	3	2	0	1	1	0	421	2, "Supernova"
[3]	0	0	34	0	0	0	0	0	0	0	0	0	34	3, "other"
[5]	0	0	0	60	0	0	0	0	1	0	0	0	61	5, "Blazar Outburst"
[6]	0	0	0	0	126	0	0	0	0	0	0	0	126	6, "AGN Variability"
[7]	0	0	0	0	0	32	0	0	0	0	0	0	32	7, "UVCeti Variable"
[8]	0	0	0	0	0	0	6	0	0	0	0	0	6	8, "Asteroid"
[9]	0	0	0	0	0	0	0	18	0	0	0	0	18	9, "Variable"
[10]	0	0	0	0	0	0	0	0	12	0	0	0	12	10, "Mira Variable"
[11]	0	0	0	0	0	0	0	0	0	43	0	0	43	11, "High Proper Motion Star"
[12]	0	0	0	0	0	0	0	0	0	0	5	0	5	12, "Comet"
[16]	0	0	0	0	0	0	0	0	0	0	0	1	1	
Total	277	405	41	61	131	33	10	23	16	47	6	2	1052	

CRTS Predictions

	[1]	[2]	[3]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[16]	Total
[1]	273	3	4	1	1	0	1	3	3	3	0	1	293
[2]	4	402	3	0	4	1	3	2	0	1	1	0	421
[3]	0	0	34	0	0	0	0	0	0	0	0	0	34
[5]	0	0	0	60	0	0	0	0	1	0	0	0	61
[6]	0	0	0	0	126	0	0	0	0	0	0	0	126
[7]	0	0	0	0	0	32	0	0	0	0	0	0	32
[8]	0	0	0	0	0	0	6	0	0	0	0	0	6
[9]	0	0	0	0	0	0	0	18	0	0	0	0	18
[10]	0	0	0	0	0	0	0	0	12	0	0	0	12
[11]	0	0	0	0	0	0	0	0	0	43	0	0	43
[12]	0	0	0	0	0	0	0	0	0	0	5	0	5
[16]	0	0	0	0	0	0	0	0	0	0	0	1	1
Total	277	405	41	61	131	33	10	23	16	47	6	2	1052

- 1, "Cataclysmic Variable"
- 2, "Supernova"
- 3, "other"
- 5, "Blazar Outburst"
- 6, "AGN Variability"
- 7, "UVCeti Variable"
- 8, "Asteroid"
- 9, "Variable"
- 10, "Mira Variable"
- 11, "High Proper Motion Star"
- 12, "Comet"

Recall → $273/277 = 98.5\%$

False Alarms → $(293-273)/293 = 7\%$

Better than saying “could be anything”

- | | Predictions 1-4 | | | | Actual | BP1 | | BP2 |
|---|-----------------|---|---|----|--------|---------|----------|---------|
| • | 1 | 2 | 0 | 11 | 2 | 69.16 % | <-Failed | 30.84 % |
| • | 1 | 2 | 0 | 11 | 2 | 50.41 % | <-Failed | 49.59 % |
| • | 1 | 2 | 0 | 11 | 2 | 50.41 % | <-Failed | 49.59 % |
| • | 2 | 7 | 6 | 1 | 7 | 99.95 % | <-Failed | 0.05 % |
| • | 2 | 1 | 0 | 6 | 1 | 72.45 % | <-Failed | 27.38 % |
| • | 2 | 1 | 0 | 11 | 1 | 50.94 % | <-Failed | 49.06 % |
| • | 2 | 1 | 0 | 11 | 11 | 98.06 % | <-Failed | 1.94 % |
| • | 2 | 6 | 1 | 0 | 9 | 63.54 % | <-Failed | 25.86 % |
| • | 2 | 1 | 0 | 6 | 1 | 96.13 % | <-Failed | 3.75 % |
| • | 2 | 6 | 1 | 11 | 6 | 63.56 % | <-Failed | 27.91 % |
| • | 6 | 2 | 1 | 0 | 1 | 33.84 % | <-Failed | 33.61 % |
| • | 1 | 2 | 0 | 11 | 5 | 60.20 % | <-Failed | 39.80 % |
| • | 1 | 2 | 0 | 6 | 2 | 50.69 % | <-Failed | 49.31 % |
| • | 1 | 2 | 0 | 6 | 2 | 50.44 % | <-Failed | 49.56 % |

<http://www.iucaa.ernet.in/~nspp>

nspp@iucaa.ernet.in

Thank You