Celestial Object Classification

Xiaoyang Wang   Ziang Zeng

Outline

1. Astronomical Challenge
2. Data & Preprocessing
3. Methodology
4. Results
5. Conclusions

Section 1

Astronomical Challenge

Astronomical Challenge

Classifying celestial objects into stars, galaxies or quasars.

Section 2

Data & Preprocessing

Images

Figure 1: Galaxy

Figure 2: Star

Figure 3: Qusar

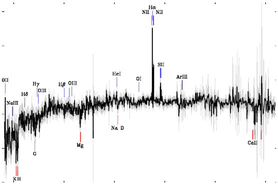


Figure 4: Galaxy Spec

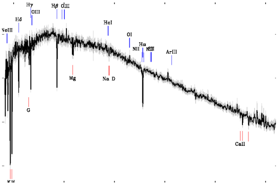


Figure 5: Star Spec

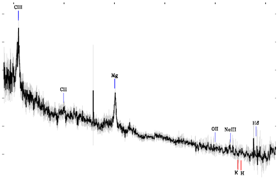


Figure 6: Qusar Spec

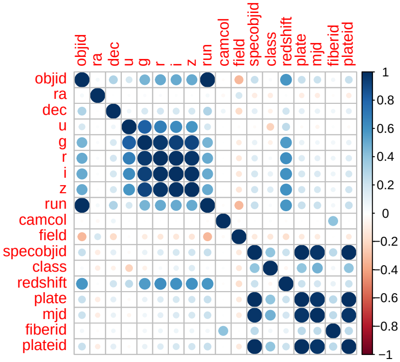
Metadata

Table 1: Metadata of the celestial objects

vars	explanations
ra	Right Ascension angle (at J2000 epoch)
dec	Declination angle (at J2000 epoch)
u	Ultraviolet filter
g	Green filter
r	Red filter
i	Near Infrared filter
z	Infrared filter
run	Run Number
rerun	Rerun Number
camcol	Camera column
field	Field number
specobjid	Unique ID used for optical spectroscopic objects
class	Object class
redshift	Redshift value based on the increase in wavelength
plate	Plate
mjd	Modified Julian Date

EDA

- Missing Values:
  - Metadata: 3, Regression Imputation
  - Image of Spectra: n
- Samples for each catagory: 33333
- Correlationship



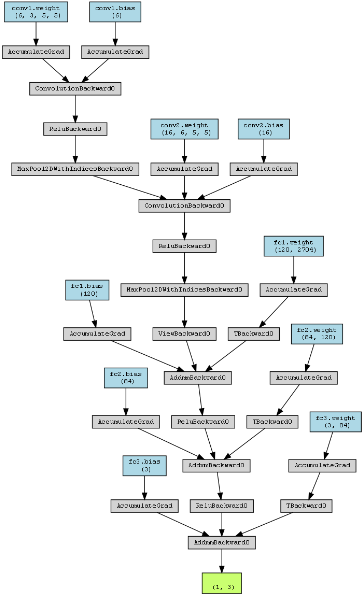
Section 3

Methodology

Meta Data

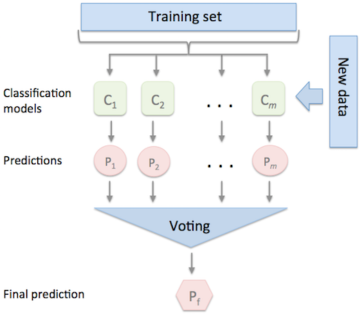
- **Explanatory Variables**: u, g, r, i, z, redshift
- **Response Variable**: class
  - GALAXY: 0
  - QSO: 1
  - STAR: 2
- **kNN**: k = 3
- **Decision Tree**:
  - Gini impurity
  - max_depth: 4
- **Logistic Regression**
  - C: 1
  - penalty: l2
  - $P(Y_i = k) = \frac{e^{\beta_k \cdot X_i}}{\sum_{j=1}^{3} e^{\beta_j \cdot X_i}},\ i = 0, 1, 2$

Images



- Structure:
  - 2 layers of convolution and 1 maxpooling
  - 3 layers of full connecting
  - output: $\vec{y} = (y_1, y_2, y_3)$
    $y_{pred} = argmax_i\{\vec{y}\}$
    Probability through softmax
    $P(y = j \mid \mathbf{z}) = \frac{e^{z_j}}{\sum_{k=1}^{3} e^{z_k}}$
- Training:
  SGD with different momentum, Adam, 10 epoch, batch size 64,lr 0.001

Voting Classifier



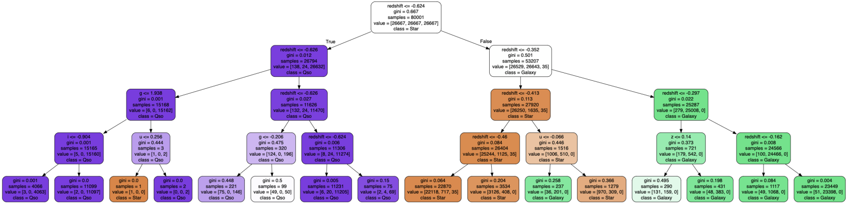
- Soft Voting:
  - Models $\{C_1, \cdots C_n\}$
  - For a given inputs, C_i has a prediction $P_i(y_j|x)$
  - The predict probabilities for voting classifier $P(y_j|x) = \frac{1}{m}\sum_{i=1}^{m} P_i(y_j|x)$
  - The prediction $p(x) = \arg\max_{y_j} P(y_j|x)$
- Hard Voting:
  - $p(x) = \text{mode}(p_1(x), p_2(x), \ldots, p_m(x))$, mode identify the most frequent one

- Construction:
  The candidate models are KNN, Logistic Regression, Classification Tree, CNN for image and CNN for spectrum image

Section 4

Results

- Decision Tree:



- Logistic Regression:

Table 2: Coefficients of Logistic Regression

	Intercept	u	g	r	i	z	redshift
Galaxy	15.09828	1.110694	-1.697660	-0.1527087	0.6143883	-0.0238249	23.35758
Qso	16.80806	-2.883331	5.212753	0.7957320	-1.2213345	-2.1409717	32.50778
Star	-31.90634	1.772637	-3.515094	-0.6430233	0.6069462	2.1647966	-55.86536

Metadata

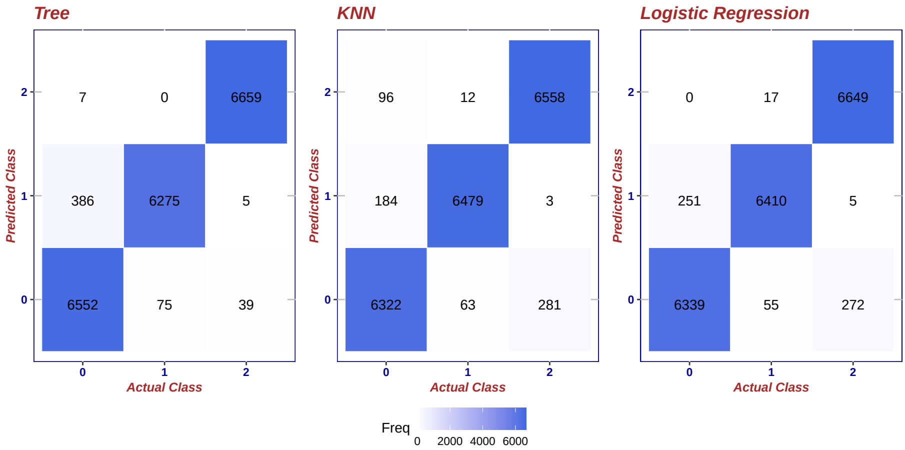


Figure 7: Confusion Matrices for Metadata Models

Images

Voting Classifier

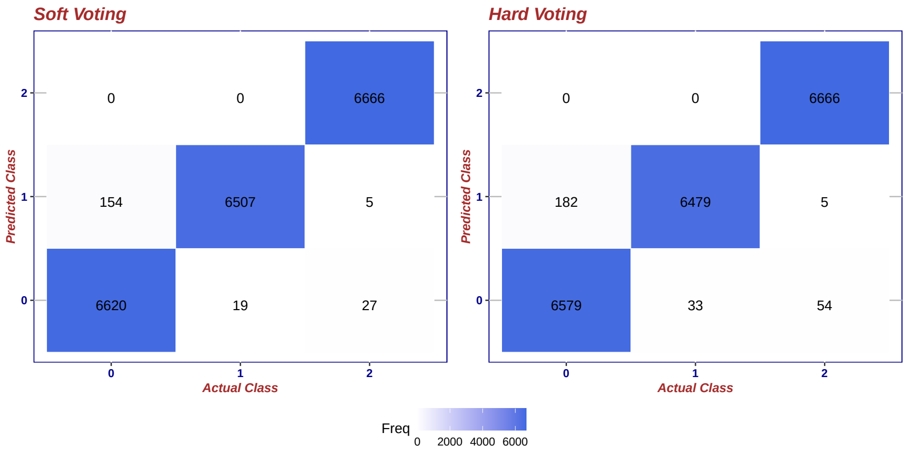


Figure 8: Confusion Matrix of Voting Classifier

Section 5

Conclusions

Conclusions

Table 3: Evaluation of Models

Data Model		M kNN	M DT	M LR	IC CNN	IS CNN	M+IC+IS VC
Accuracy		96.79%	97.68%	97.04%	93.91%	99.14%	97.8%
Precision	Star	95.59%	99.82%	95.95%	0%	0%	0%
	Galaxy	96.01%	96.45%	96.46%	0%	0%	0%
	Qso	98.84%	96.8%	98.77%	0%	0%	0%
Recall	Star	98.67%	99.82%	99.59%	0%	0%	0%
	Galaxy	94.61%	96.68%	95.06%	0%	0%	0%
	Qso	97.13%	96.56%	96.5%	0%	0%	0%
F1	Star	97.11%	99.82%	97.74%	0%	0%	0%
	Galaxy	95.31%	96.56%	95.76%	0%	0%	0%
	Qso	97.98%	96.68%	97.63%	0%	0%	0%

References I

[1] Jialin Gao, Jianyu Chen, Jiaqi Wei, Bin Jiang, and A-Li Luo. Deep multimodal networks for m-type star classification with paired spectrum and photometric image. *Publications of the Astronomical Society of the Pacific*, 135:044503, 05 2023.