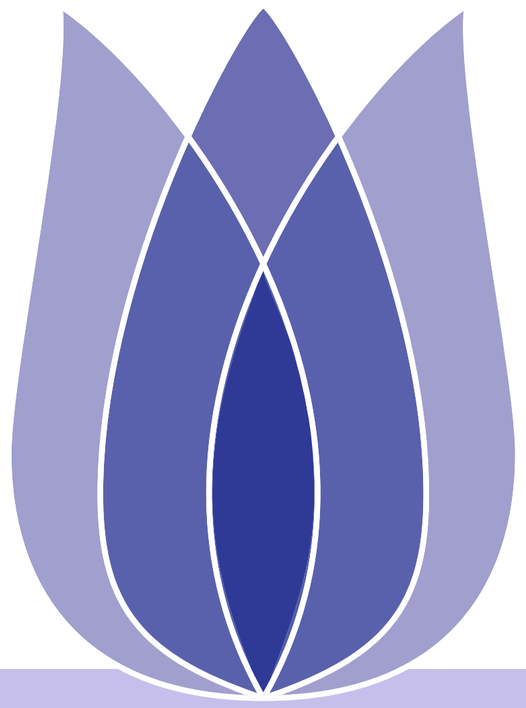
FLIP(01) mid-term Presentation

Xiaojuan Cheng
Xi'an Shiyou University

October 22, 2019



©*FLIP(01)*

Outline

Introduction

Data Description

Features Distribution

Model Comparisions

Conclusion

Introduction

Data Description

Features Distribution

Model Comparisions

Conclusion

Introduction
Problem Description
Data Description
Features Distribution
Model Comparisions
Conclusion

Introduction

Problem Description

Introduction
Problem Description
Data Description
Features Distribution
Model Comparisions
Conclusion

- This is a classification problem to predict seed varieties. There are 210 samples, consistent of 7 features and 1 label. The following is some requirements.

Problem Description

Introduction
Problem Description
Data Description
Features Distribution
Model Comparisions
Conclusion

- This is a classification problem to predict seed varieties. There are 210 samples, consistent of 7 features and 1 label. The following is some requirements.
  - According to the given train data set, training a classified model and then using the model to predict the class of Wheat Seeds.
  - Estimated the error of the training classification model.

Introduction

Data Description

Data Description

Data Description

Features Distribution

Model Comparisions

Conclusion

Data Description

Data Description

Introduction

Data Description

Data Description

Data Description

Features Distribution

Model Comparisions

Conclusion

- Attribute Information

1. There are 8 attributes, including 1 class attribute and 7 feature attributes.
2. The detailed description of the data is shown in the following table.

Data Description

Introduction

Data Description

Data Description

Data Description

Features Distribution

Model Comparisions

Conclusion

	area	peri	compactness	glength	gwidth	Asymmetry coefficient	Graft length of the grain	label
0	14.88	14.57	0.8811	5.554	3.333	1.0180	4.956	1
1	14.29	14.09	0.9050	5.291	3.337	2.6990	4.825	1
2	13.84	13.94	0.8955	5.324	3.379	2.2590	4.805	1
3	16.14	14.99	0.9034	5.658	3.562	1.3550	5.175	1
4	14.38	14.21	0.8951	5.386	3.312	2.4620	4.956	1
5	14.69	14.49	0.8799	5.563	3.259	3.5860	5.219	1
6	14.11	14.10	0.8911	5.420	3.302	2.7000	5.000	1
7	16.63	15.46	0.8747	6.053	3.465	2.0400	5.877	1
8	16.44	15.25	0.8880	5.884	3.505	1.9690	5.533	1

Figure 1: Data

Introduction

Data Description

Features Distribution

Features Distribution

Model Comparisions

Conclusion

Features Distribution

Features Distribution

Introduction

Data Description

Features Distribution

Features Distribution

Model Comparisions

Conclusion

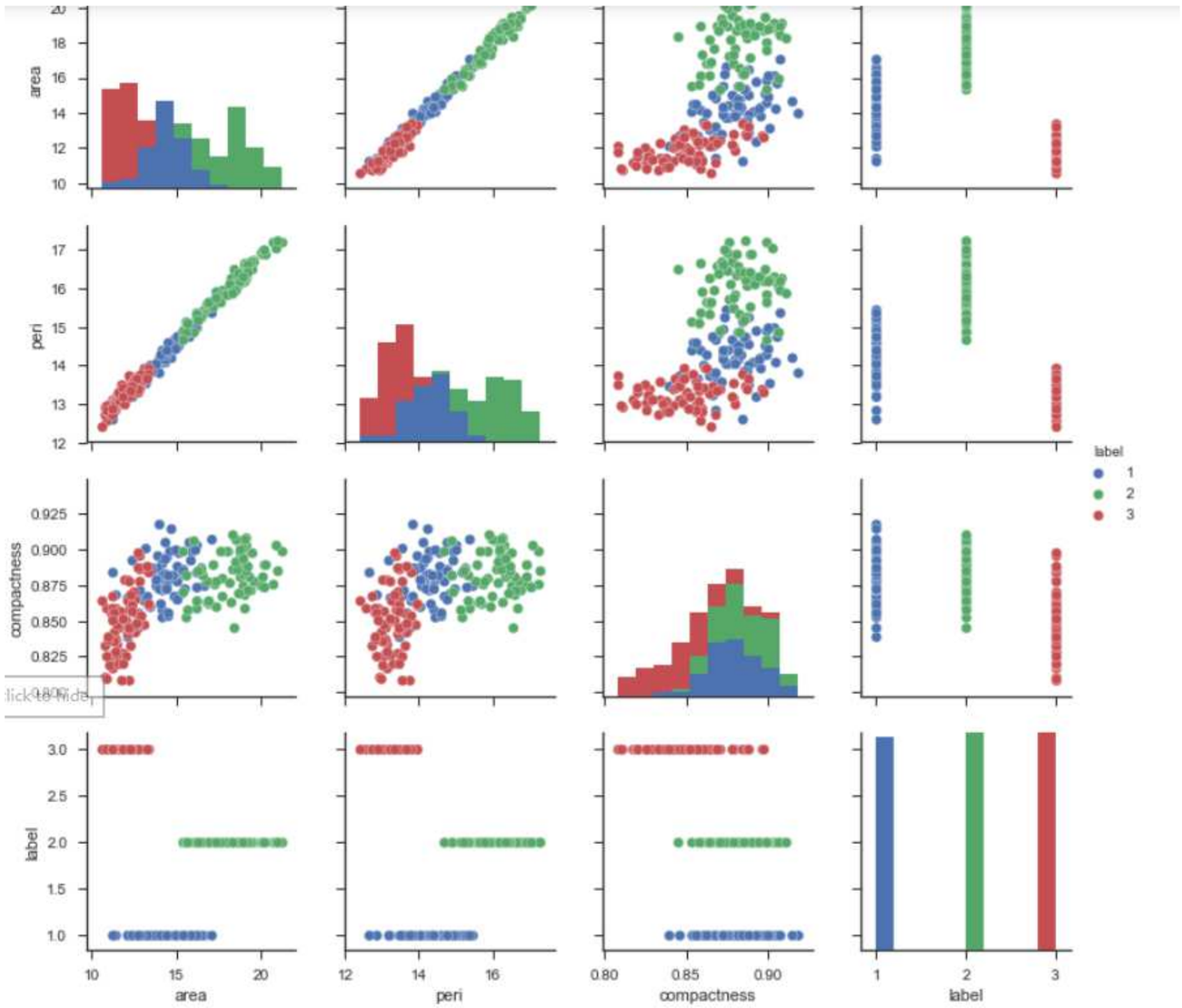


Figure 2: The distribution of features

Introduction

Data Description

Features Distribution

Model Comparisions

Model Comparisions

Conclusion

Model Comparisions

Model Comparisions

Introduction
Data Description
Features Distribution
Model Comparisions
Model Comparisions
Conclusion

KNN performance:

	precision	recall	f1-score	support
class1	0.88	0.82	0.85	17
class2	1.00	0.95	0.98	21
class3	0.82	0.93	0.87	15
avg / total	0.91	0.91	0.91	53

Randomforest performance:

	precision	recall	f1-score	support
class1	0.86	0.71	0.77	17
class2	0.95	0.95	0.95	21
class3	0.78	0.93	0.85	15
avg / total	0.87	0.87	0.87	53

Figure 3: The Comparision of two models

Introduction

Data Description

Features Distribution

Model Comparisions

Conclusion

Summary

Conclusion

Summary

Introduction

Data Description

Features Distribution

Model Comparisions

Conclusion

Summary

- From the above result presentation, we can find that
  KNN is better than Randomforest model.

- From the feature distribution, we can find that
  Class1's data distribution is more nearly Gaussian distribution.
  The first two features are more interrelated.

Thank you & Question