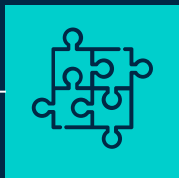


Data & Analytics – Iris Flower

Yang Leyuan
National University of Singapore
Business Analytics & Statistics

TABLE OF CONTENTS



01

Problem & Dataset

- Motivation
- Data Quality
- Data Transformation



02

Data Modelling

- Logistics Regression
- KNN Classifier
- Support Vector Classifier



03

Visualization

- Classification & Similarity
- Limitation

Motivation

Retrieve records of
the most similar Iris
flowers in his garden
for any input Iris
flower based on
historical data

→
Classify the input data
belongs to which class
then find nearest
neighbors

-Analyse collected data

- 1) Features
- 2) Correlations
- 3) Outliers

-Model selection & Training (Multiclass model)

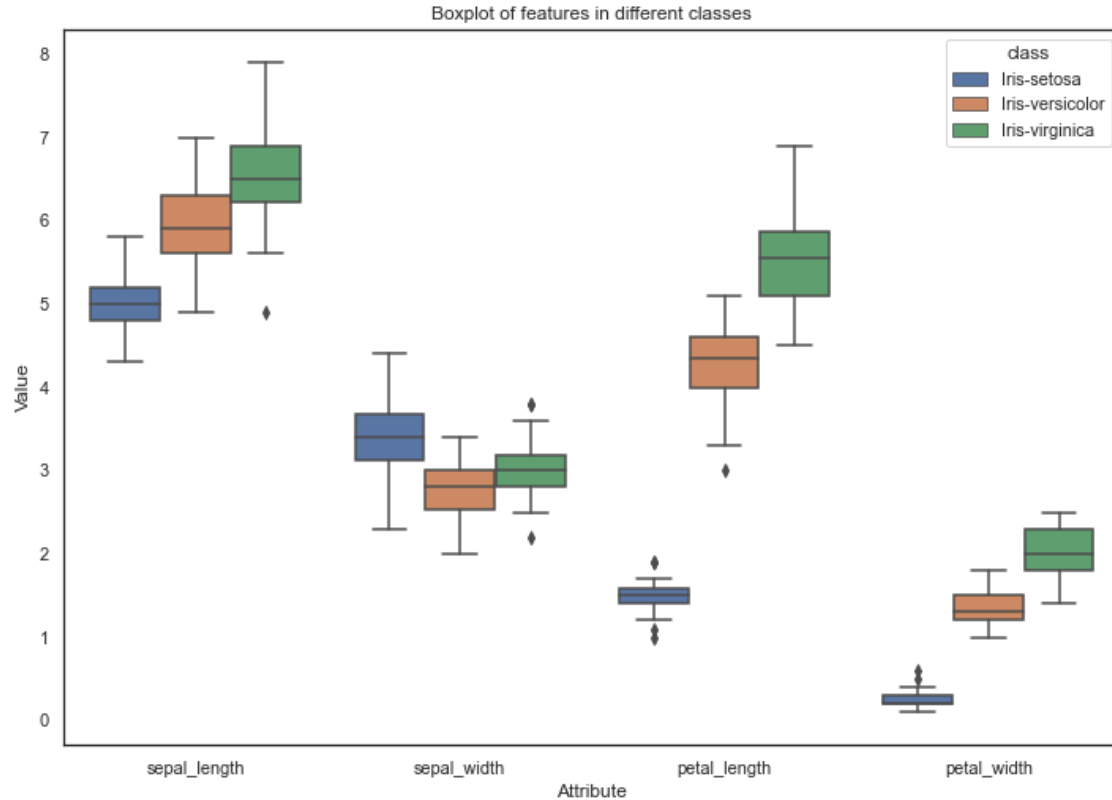
- 1) Logistics Regression
- 2) K-nearest neighbours Classifier
- 3) Support Vector Classifier

-Prediction

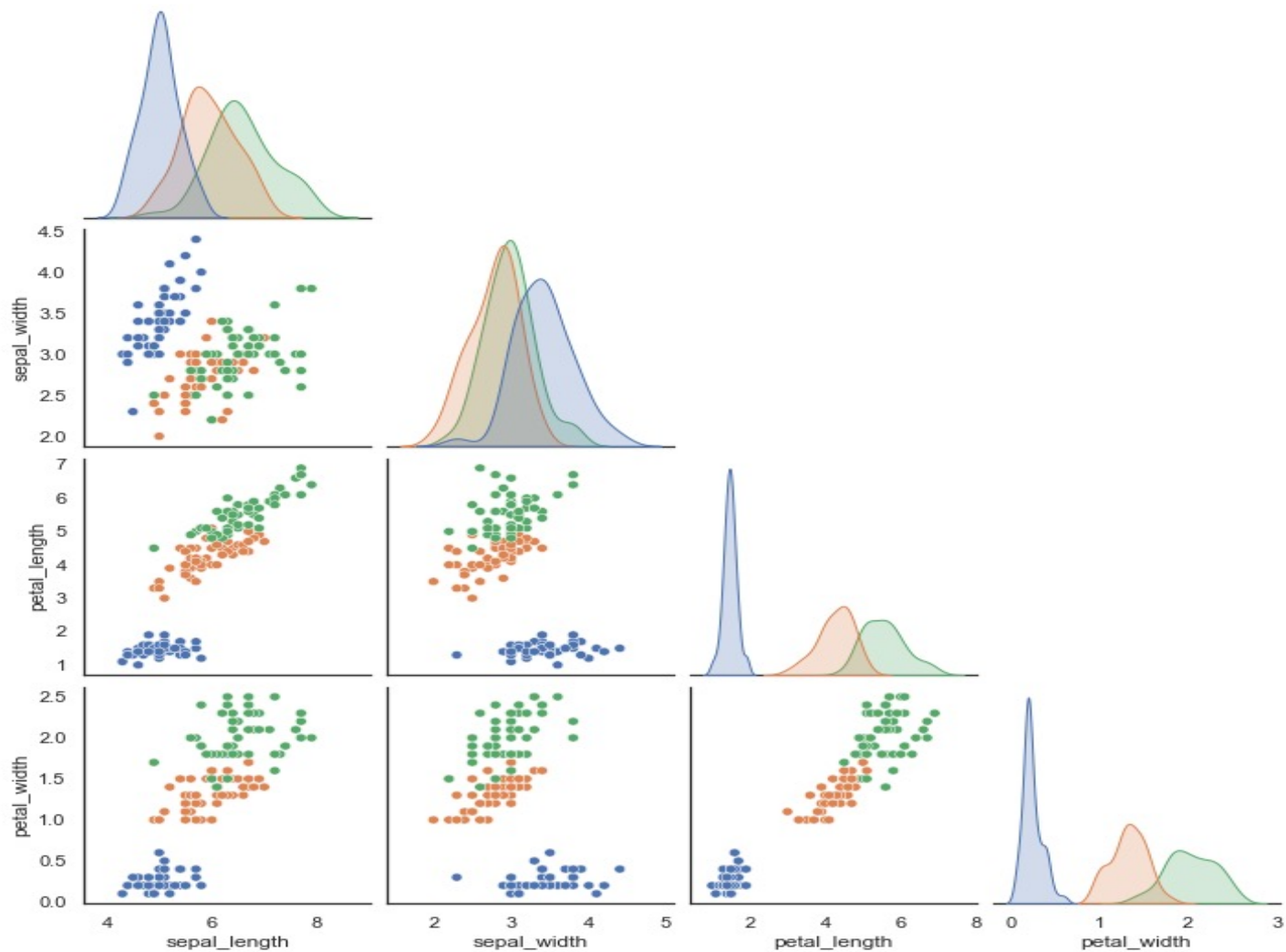
Data Quality

The background is a solid dark blue. It features several abstract geometric elements: a teal square on the left, a small teal square at the bottom right, a small pink square at the top right, and a small orange square on the right side. There are also several thin white lines of varying lengths and orientations, including a vertical line on the left, a horizontal line at the bottom, and a diagonal line on the right.

Correlation



Abnormal sepal
width data of Setosa
class



Conclusion of data quality

- Sample size too small
- Features are highly correlated
- Abnormality of records



Data Modelling

The background is a solid dark blue. It features several abstract geometric elements: a teal square at the bottom left, a teal square at the bottom right, a teal square at the top left, a teal square at the top right, a pink square at the top center, a pink square at the top right, a pink square at the bottom right, and a pink square at the bottom center. There are also several thin white lines of varying lengths and orientations scattered across the background.

Split data

Training data

70%

Testing data

30%

~~20%?~~
Accuracy is 1.0



Logistics Regression

LR:		precision	recall	f1-score	support
	0	1.00	1.00	1.00	16
	1	1.00	0.94	0.97	18
	2	0.92	1.00	0.96	11
	accuracy			0.98	45
	macro avg	0.97	0.98	0.98	45
	weighted avg	0.98	0.98	0.98	45

LR_ROC_AUC_SCORE: 0.9968457266932212

KNN

KN:		precision	recall	f1-score	support
	0	1.00	1.00	1.00	16
	1	1.00	0.94	0.97	18
	2	0.92	1.00	0.96	11
	accuracy			0.98	45
	macro avg	0.97	0.98	0.98	45
	weighted avg	0.98	0.98	0.98	45

KN_ROC_AUC_SCORE: 0.9842616433030376

SVC



SVC:		precision	recall	f1-score	support
	0	1.00	1.00	1.00	16
	1	1.00	0.94	0.97	18
	2	0.92	1.00	0.96	11
	accuracy			0.98	45
	macro avg	0.97	0.98	0.98	45
	weighted avg	0.98	0.98	0.98	45

SVC_ROC_AUC_SCORE: 0.9968457266932212

Data Visualization

The background is a solid dark blue. It features several abstract geometric elements: a teal square on the left, a small teal square at the bottom right, a small pink square at the top right, a small orange square below it, and a small teal square further down. There are also several thin white lines: a vertical line on the left, a vertical line near the top center, a vertical line on the right, and a horizontal line at the bottom that is teal on the left and pink on the right.

Similar Points within the class

Sepal Length:

Sepal Width:

Petal Length:

Petal Width:

sepal_length	sepal_width	petal_length	petal_width	class	label
4	6	1	2	Iris-setosa	0
5.7	4.4	1.5	0.4	Iris-setosa	0
5.5	4.2	1.4	0.2	Iris-setosa	0
5.2	4.1	1.5	0.1	Iris-setosa	0
5.4	3.9	1.3	0.4	Iris-setosa	0
5.1	3.8	1.5	0.3	Iris-setosa	0
5.1	3.7	1.5	0.4	Iris-setosa	0
4.6	3.6	1	0.2	Iris-setosa	0
5.1	3.8	1.9	0.4	Iris-setosa	0
5.4	3.9	1.7	0.4	Iris-setosa	0
5	3.5	1.6	0.6	Iris-setosa	0





Limitations

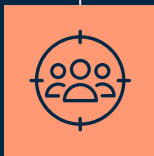
Over/Underfitting

Cross validation



Interface

UI design
Various plots



Other ML models

DecisionTree Classifier
XGBoost
Random Forest
.....

Accuracy

Split Petal & Sepal into
two datasets and build
models for each

The background is a dark blue field decorated with a pattern of small, colorful squares (pink, orange, and teal) and thin white vertical lines of varying lengths, creating a modern, minimalist aesthetic.

Questions?

Thank you!