

Iris Data Set Classification

BY GROUP CLASSIFICATION III





Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Search

☒ Repository ☐ Web

Google™

[View ALL Data Sets](#)

Iris Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Famous database; from Fisher, 1936



Data Set Characteristics:	Multivariate	Number of Instances:	150	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	4	Date Donated	1988-07-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	3184160

IRIS DATA SET

<https://archive.ics.uci.edu/ml/datasets/Iris>



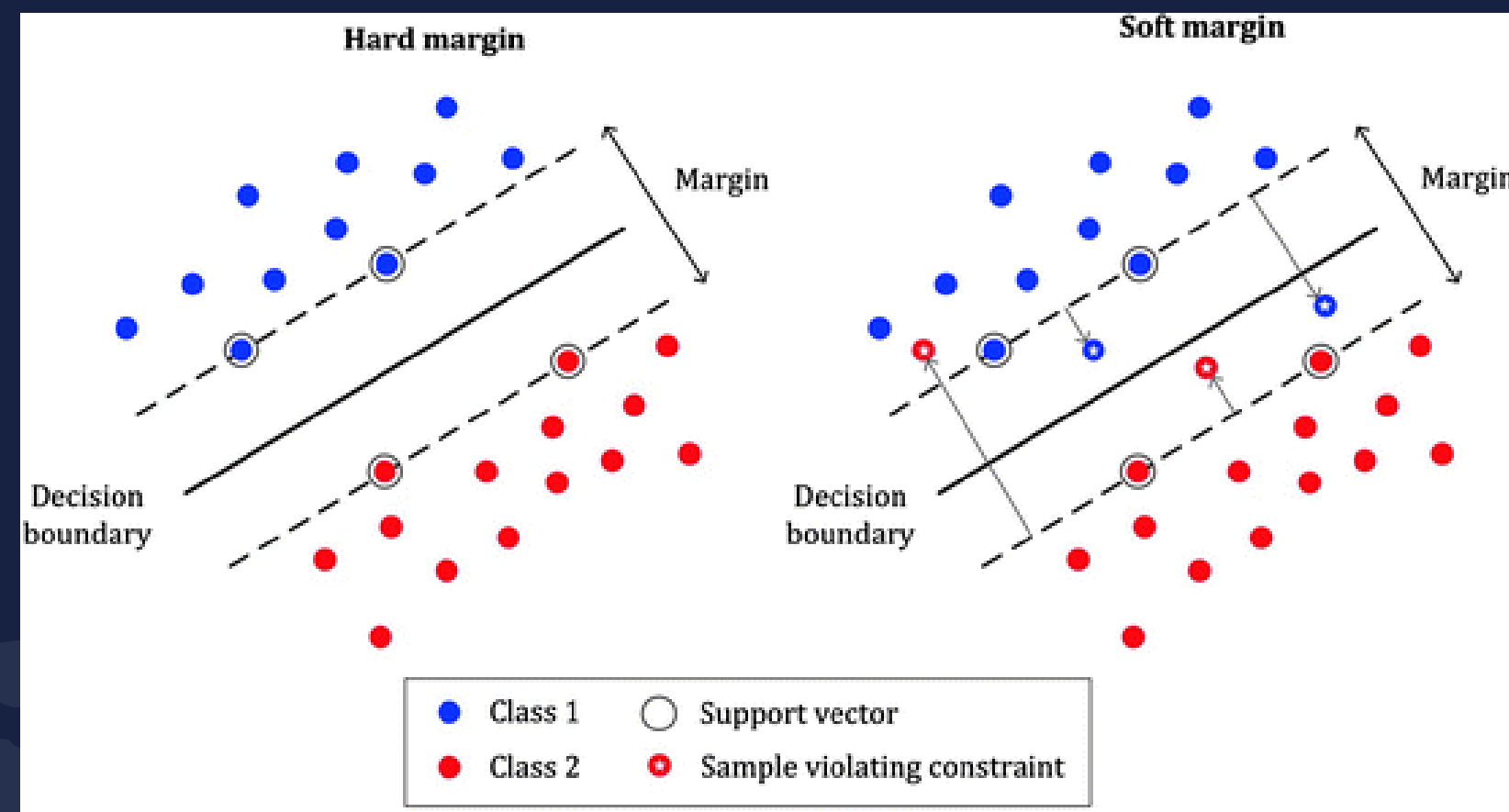
Incanter Dataset				
Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa
5.4	3.4	1.7	0.2	setosa
5.1	3.7	1.5	0.4	setosa

- **DATA NAME**
- **ATTRIBUTES NUMBER : 4**
- **ATTRIBUTES NAME :**
sepal length, sepal width, petal length, petal width
- **ATTRIBUTES DETAIL :**
1.sepal length in cm
2.sepal width in cm
3.petal length in cm
4.petal width in cm
5.class - (setosa , versicolour , virginica)
- **CLASS NAME :**
iris Setosa(50) , iris Versicolour(50) , iris Virginica(50)
- **NUMBER OF INSTANCES : 150**

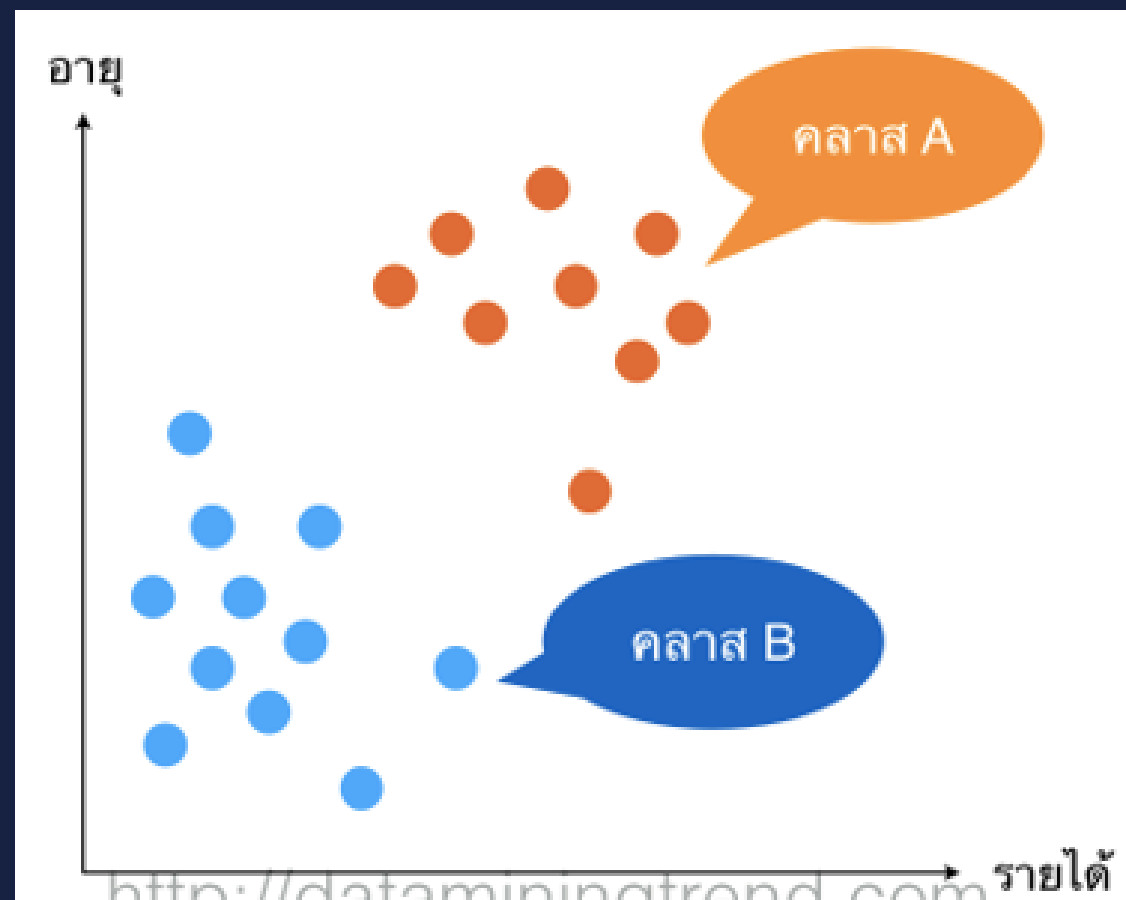
TECHNIQUE

- NAME : SVM(SUPPORT VECTOR MACHINE)

เป็นอัลกอริทึมที่ใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูล โดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกแยกกลุ่มข้อมูลได้ดีที่สุด

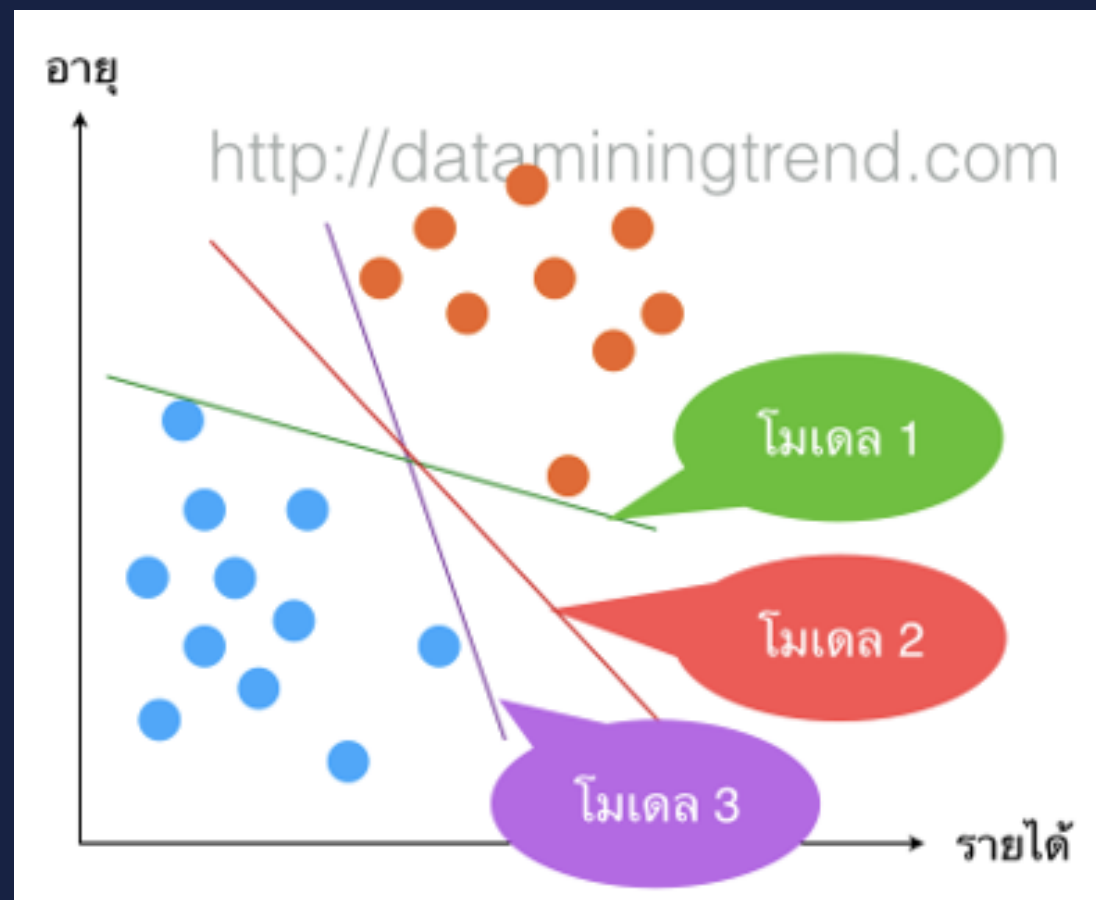


TECHNIQUE : SVM



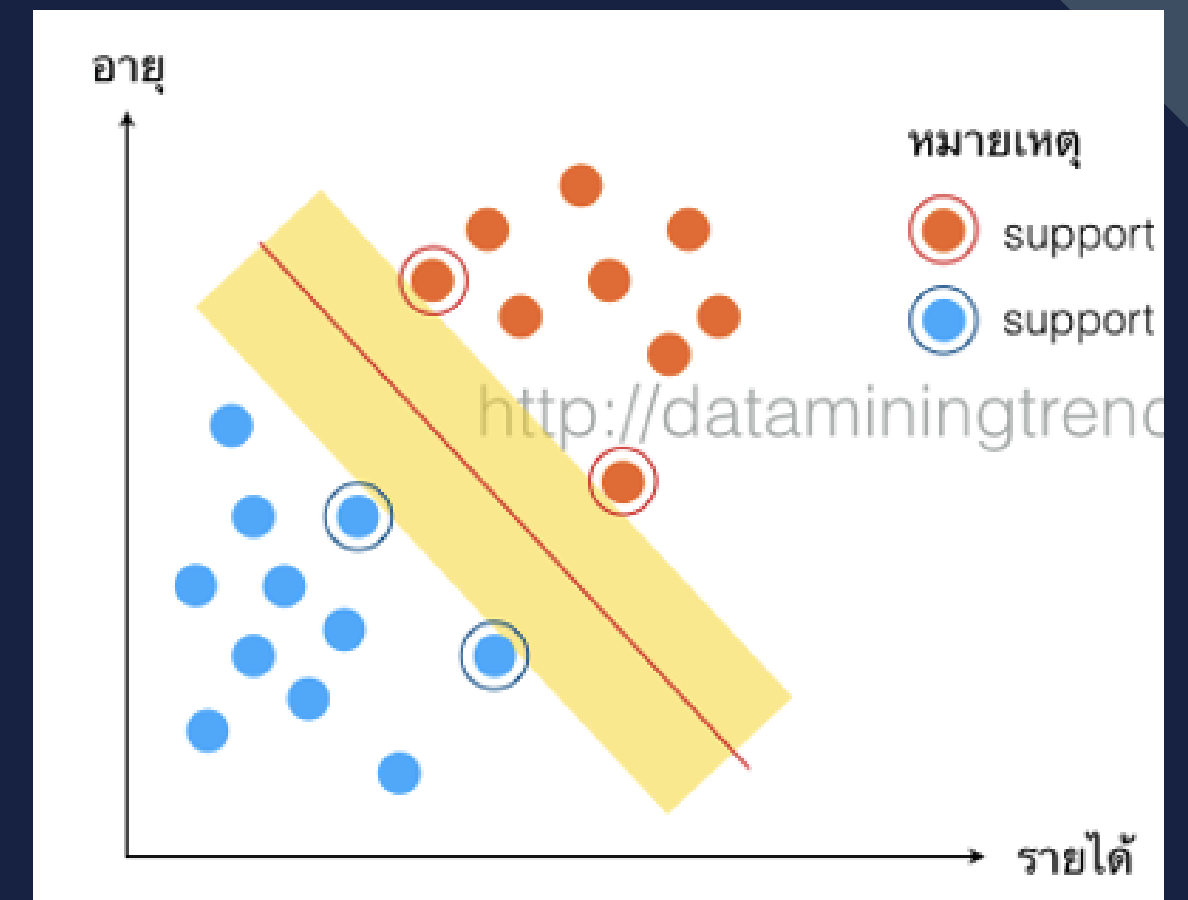
STEP1 :

วางข้อมูลลงในฟีเจอร์สเปซ
(Feature Space)



STEP2 :

สร้างเส้นแบ่ง (Hyperplane) ที่เป็น
เส้นตรงขึ้นมาเพื่อแบ่งสองกลุ่มออก
จากกัน



STEP3 :

เลือก linear model ที่มีระยะห่าง
ระหว่าง 2 คลาสห่างกันมากที่สุด ดัง
แสดงในรูป

COMMAND NAME OF TECHNIQUE

- USED IN R PROGRAM

```
> data_iris <- iris
> str(data_iris)
'data.frame': 150 obs. of 5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
> summary(data_iris)
 Sepal.Length Sepal.Width Petal.Length Petal.Width Species
Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100 setosa :50
1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300 versicolor:50
Median :5.800 Median :3.000 Median :4.350 Median :1.300 virginica :50
Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199
3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800
Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500
```

COMMAND NAME OF TECHNIQUE

- USED IN R PROGRAM

```
> set.seed(123)
> install.packages("caTools")
> library(caTools)
> data_split <- sample.split(data_iris$Species, SplitRatio = 0.7)
> training_set <- subset(data_iris, data_split == TRUE)
> test_set <- subset(data_iris, data_split == FALSE)
> training_set[, 1:4] = scale(training_set[, 1:4])
> test_set[, 1:4] = scale(test_set[, 1:4])
> install.packages("e1071")
> library(e1071)
> mymodel <- svm(Species ~ ., data = iris)
> classifier1 = svm(formula = Species ~ ., data = training_set, type = 'C-classification', kernel = 'radial')
> classifier2 = svm(formula = Species ~ Petal.Width + Petal.Length, data = training_set,
+ type = 'C-classification', kernel = 'radial')
```

COMMAND NAME OF TECHNIQUE

- USED IN R PROGRAM

```
> test_pred1 = predict(classifier1, type = 'response', newdata = test_set[-5])  
> test_pred2 = predict(classifier2, type = 'response', newdata = test_set[-5])  
> cm1 = table(test_set[,5], test_pred1)  
> cm2 = table(test_set[,5], test_pred2)  
> ACC <- sum( diag (cm1) ) / nrow ( test_set )
```


PERFORMANCE MEASUREMENT

- CONFUSION MATRIX OF "TEST DATA (IRIS)"

		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

True Position of Model (TP) = $15+13+14 = 42$

- **CONFUSION MATRIX OF “SETOSA”**

TP = 15		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

TN = 15		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

FP = 0		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

FN = 0		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

Precision = $15 / (15 + 0 + 0) = 1$

Recall = $15 / (15 + 0 + 0) = 1$

- **CONFUSION MATRIX OF “VERSICOLOR”**

TP = 13

		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

TN = 16

		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

FP = 1

		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

FN = 2

		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

Precision = $13 / (13 + 1 + 0) = 0.929$

Recall = $13 / (13 + 2 + 0) = 0.867$

- **CONFUSION MATRIX OF “VIRGINICA”**

TP = 14

		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

TN = 17

		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

FP = 2

		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

FN = 1

		Prediction		
		setosa	versicolor	virginica
Actual	setosa	15	0	0
	versicolor	0	13	2
	virginica	0	1	14

Precision = $14 / (14 + 2 + 0) = 0.875$

Recall = $14 / (14 + 1 + 0) = 0.933$



Thank You

GROUP CLASSIFICATION III

MEMBER :

07610477 นายศักดิ์ณรงค์ สมบัติเจริญ

620710405 นางสาวณัฐริดา ลากธนชัย

620710407 นางสาวเพชรรัตน์ สุขอุบล

620710408 นางสาวสุราทิพย์ แยมกลั่น