


```
from numpy import matrix
```

```
>>> A = matrix( [ [0,7,3], [12,11,35], [91,22,52] ] )
```

```
>>> A
```

```
matrix([[ 1,  2,  3],  
        [11, 12, 13],  
        [21, 22, 23]])
```

```
>>> B = matrix( [ [ 4, 8, 9] ] )
```

```
>>> B
```

```
matrix([[4, 8, 9]])
```

```
>>> C = matrix([ [7], [8], [5] ] )
```

```
>>> C
```

```
matrix([[7],  
        [8],  
        [5]])
```

```
from numpy import matrix
```

```
from numpy import linalg
```

```
>>> A = matrix( [ [0,7,3], [12,11,35], [91,22,52] ] )
```

```
# Transpose
```

```
>>> A.T
```

```
matrix([[ 0, 12, 91],  
        [ 7, 11, 22],  
        [ 3, 35, 52]])
```

```
# Matrix multiplication
```

```
>>> A * C
```

```
# Inverse of A
```

```
>>> A.I
```

```
# Solving linear equation system
```

```
>>> linalg.solve(A, x)
```

◆ NumPy

- For scientific computing
- Functionality for multi-dimensional arrays, high-level mathematical functions
- ndarray (n-dimensional) class is the core functionality

◆ SciPy

- Collection of functions for scientific computing (advanced linear algebra, statistical distributions, mathematical function optimization, etc.)
- `scipy.sparse`

◆ scikit-learn

- It uses NumPy arrays as data structure
- It draws from scipy functions for implementing its algorithms
- Data has to be converted to NumPy array

◆ **matplotlib**

- Primary scientific plotting library
- Functions for data visualizations such as histograms, scatter diagram, line charts, pie charts, etc.

◆ **pandas**

- For data wrangling and analysis
- Built around DataFrame (same as in R)
- It has ability to ingest from a varieties of file formats and databases, such as SQL, Exel, CSV, etc.

Example: Plotting Sin(x)

```
import numpy as np
import matplotlib.pyplot as plt

x = np.linspace(-20, 20, 100)
print(x)
# Create a second array using sine
y = np.sin(x)
# The plot function makes a line chart of one array against another
plt.plot(x, y, color="blue", marker="x", linestyle='--')

plt.show()
```

Markers

character	description
'.'	point marker
','	pixel marker
'o'	circle marker
'v'	triangle_down marker
'^'	triangle_up marker
'<'	triangle_left marker
'>'	triangle_right marker
'1'	tri_down marker
'2'	tri_up marker
'3'	tri_left marker
'4'	tri_right marker
's'	square marker
'p'	pentagon marker
'*'	star marker
'h'	hexagon1 marker
'H'	hexagon2 marker
'+'	plus marker
'x'	x marker
'D'	diamond marker
'd'	thin_diamond marker
' '	vline marker
'_'	hline marker

Line Styles

character	description
'-'	solid line style
'--'	dashed line style
'-.'	dash-dot line style
':'	dotted line style

Colors

The supported color abbreviations are the single letter codes

character	color
'b'	blue
'g'	green
'r'	red
'c'	cyan
'm'	magenta
'y'	yellow
'k'	black
'w'	white

- ◆ A 2-D labelled data structure with columns of potentially different types

Columns

	Age	Location	Name
Index 0	37	Mumbai	Shiva
1	28	Delhi	Ankit
2	39	Raipur	Prashant
3	42	Raipur	Hemant
4	36	Kolkata	Abhinav
5	31	Bangluru	Deepak

- ◆ A 2-D labelled data structure with columns of potentially different types

```
from pandas import DataFrame
```

Data frame from dictionary

```
dic = {"performance": [4, 1, 3, 4, 5, 5, 6, 7],  
      "custSatisfaction": [5, 2, 4, 5, 2, 4, 2, 7],  
      "responsiveness": [4, 2, 5, 7, 4, 3, 6, 6] }
```

```
df = DataFrame(dic, index = [1, 2, 3, 4, 5, 6, 7, 8])
```

- ◆ A 2-D labelled data structure with columns of potentially different types

```
from pandas import DataFrame
```

Accessing column

```
df["custSatisfaction"]
```

Adding column

```
df["product"] = ["MOP", "GHT", "ASD", "BHN", "ASE", "EFG", "TGH", "HUK"]
```

Dropping column

```
del df["responsiveness"]
```

OR

```
df.pop("responsiveness")
```

Machine Learning



Features

Target

Org_type	Partner	Age	Designation	Supplier_gc	Customer_gc	Motivation_1	Operation_type
Manufacturer	Manufacturer	between	Manager	Yes	Yes	4	Automobile industry
Manufacturer	Manufacturer	between	Manager	Yes	Yes	4	Automobile industry
Manufacturer	Manufacturer	between	Others	Yes	Yes	4	Food
Manufacturer	Manufacturer; Se	<=30	Others	Yes	Yes	4	Automobile industry
Manufacturer	Manufacturer; Se	<=30	Others	Yes	Yes	4	Automobile industry
Manufacturer	Manufacturer; Se	50 and ab	Vice President/ P	Yes	Yes	5	Pharmaceutical Industry
Manufacturer	Service Provider;	between	Manager	Yes	Yes	2	Electricals
Service Provider	Service Provider	between	Others	No	Yes	4	Others
Manufacturer	Manufacturer	<=30	Others	No	No	3	Automobile industry
Manufacturer	Manufacturer	between	Manager	Yes	Yes	4	Automobile industry
Manufacturer	Manufacturer	<=30	Manager	No	No	4	Metals
Manufacturer	Manufacturer	<=30	Consultant/Analy	Yes	Yes	5	Food
Service Provider	Manufacturer; Se	<=30	Director	No	No	4	Others
Service Provider	Manufacturer; Re	between	Manager	Yes	No	4	Electricals
Service Provider	Manufacturer; Se	between	Others	No	No	4	Others
Service Provider	Service Provider	<=30	Others	Yes	No	2	Automobile industry
Wholesaler/Distr	Manufacturer; Re	50 and ab	Manager	Yes	Yes	5	Others
Manufacturer	Manufacturer; Se	<=30	Manager	Yes	Yes	4	Automobile industry
Manufacturer	Manufacturer; Wf	<=30	Others	Yes	Yes	2	Automobile industry
Service Provider	Service Provider;	<=30	Others	Yes	Yes	5	Software Industry
Manufacturer	Manufacturer; Se	between	Manager	Yes	Yes	3	Automobile industry
Service Provider	Manufacturer; Wf	50 and ab	Others	No	Yes	4	Software Industry
Wholesaler/Distr	Manufacturer; Se	<=30	Others	Yes	No	2.089715758	Others
Manufacturer	Manufacturer		Manager	No	Yes	3	Others

X

y

Machine Learning



Features

Target

Org_type	Partner	Age	Designation	Supplier_gc	Customer_gc	Motivation_1	Operation_type
Manufacturer	Manufacturer	between	Manager	Yes	Yes	4	Automobile industry
Manufacturer	Manufacturer	between	Manager	Yes	Yes	4	Automobile industry
Manufacturer	Manufacturer	between	Others	Yes	Yes	4	Food
Manufacturer	Manufacturer; Se	<=30	Others	Yes	Yes	4	Automobile industry
Manufacturer	Manufacturer; Se	<=30	Others	Yes	Yes	4	Automobile industry
Manufacturer	Manufacturer; Se	50 and ab	Vice President/ P	Yes	Yes	5	Pharmaceutical Industry
Manufacturer	Service Provider;	between	Manager	Yes	Yes	2	Electricals
Service Provider	Service Provider	between	Others	No	Yes	4	Others
Manufacturer	Manufacturer	<=30	Others	No	No	3	Automobile industry
Manufacturer	Manufacturer	between	Manager	Yes	Yes	4	Automobile industry
Manufacturer	Manufacturer	<=30	Manager	No	No	4	Metals
Manufacturer	Manufacturer	<=30	Consultant/Analy	Yes	Yes	5	Food
Service Provider	Manufacturer; Se	<=30	Director	No	No	4	Others
Service Provider	Manufacturer; Re	between	Manager	Yes	No	4	Electricals
Service Provider	Manufacturer; Se	between	Others	No	No	4	Others
Service Provider	Service Provider	<=30	Others	Yes	No	2	Automobile industry
Wholesaler/Distri	Manufacturer; Re	50 and ab	Manager	Yes	Yes	5	Others
Manufacturer	Manufacturer; Se	<=30	Manager	Yes	Yes	4	Automobile industry
Manufacturer	Manufacturer; Wf	<=30	Others	Yes	Yes	2	Automobile industry
Service Provider	Service Provider;	<=30	Others	Yes	Yes	5	Software Industry
Manufacturer	Manufacturer; Se	between	Manager	Yes	Yes	3	Automobile industry
Service Provider	Manufacturer; Wf	50 and ab	Others	No	Yes	4	Software Industry
Wholesaler/Distri	Manufacturer; Se	<=30	Others	Yes	No	2.089715758	Others
Manufacturer	Manufacturer		Manager	No	Yes	3	Others

X_train

X_test

y_train

y_test

Features = dimension = attributes of the data

Class = target = category

- ◆ Machine Learning is the study of algorithms that
 - improve their performance P
 - at some task T
 - with experience E .

A well-defined learning task is given by $\langle P, T, E \rangle$

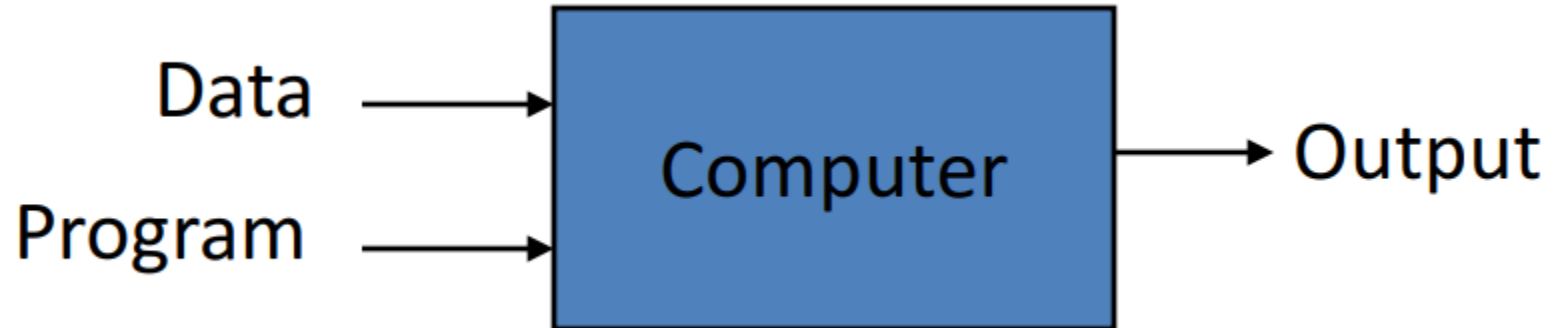
- Tom Mitchell (1998)

- ◆ The ability of a machine to improve its performance based on previous results.

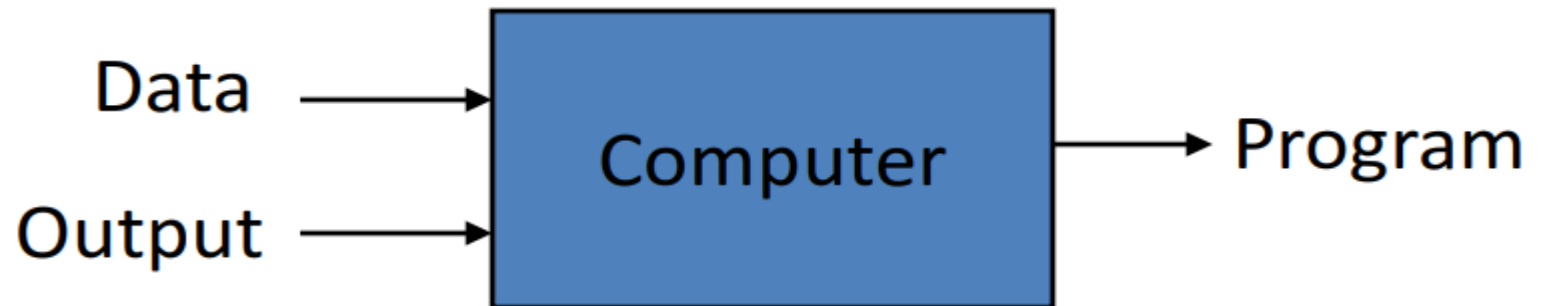
- Dictionary.com



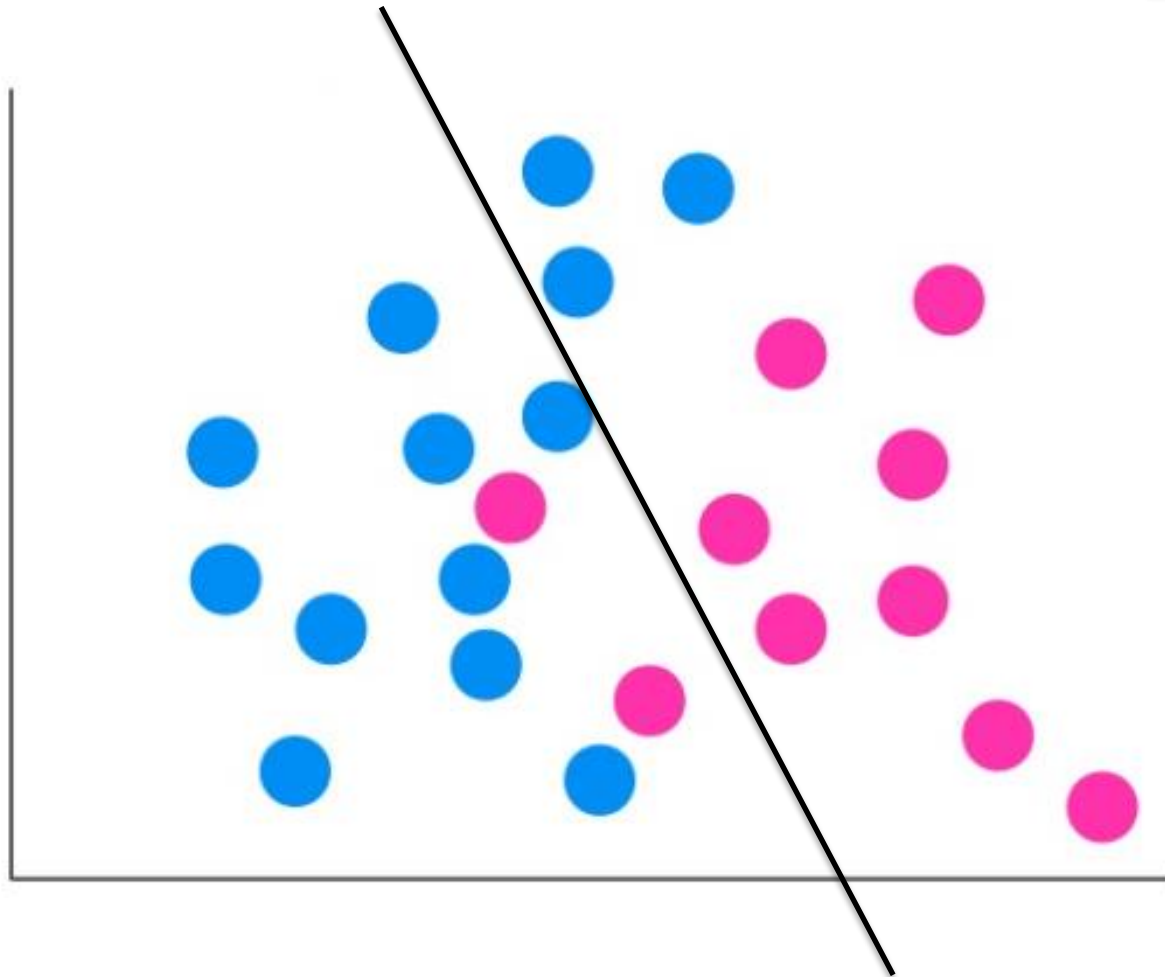
Traditional Programming



Machine Learning

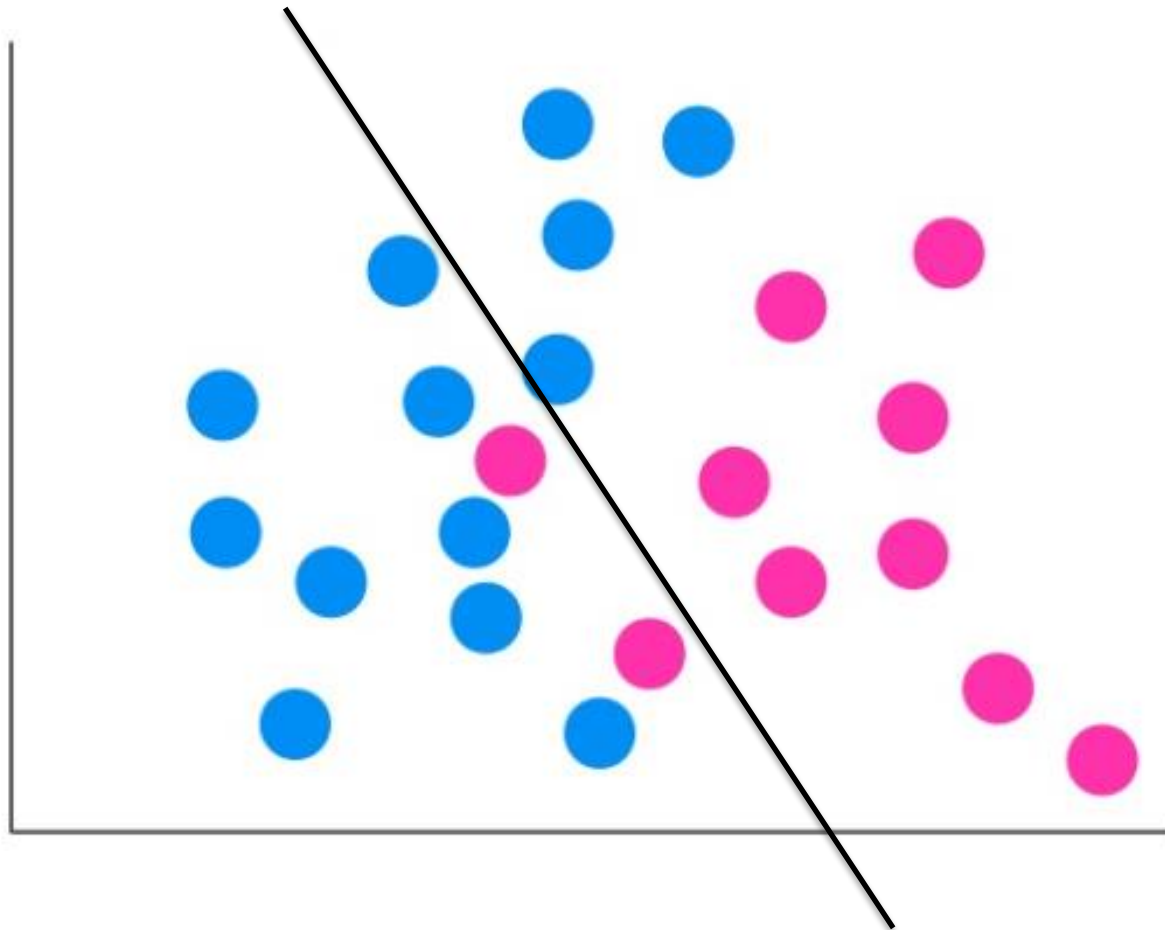


Classification



3 blue dots and 2 red dots are wrongly classified.

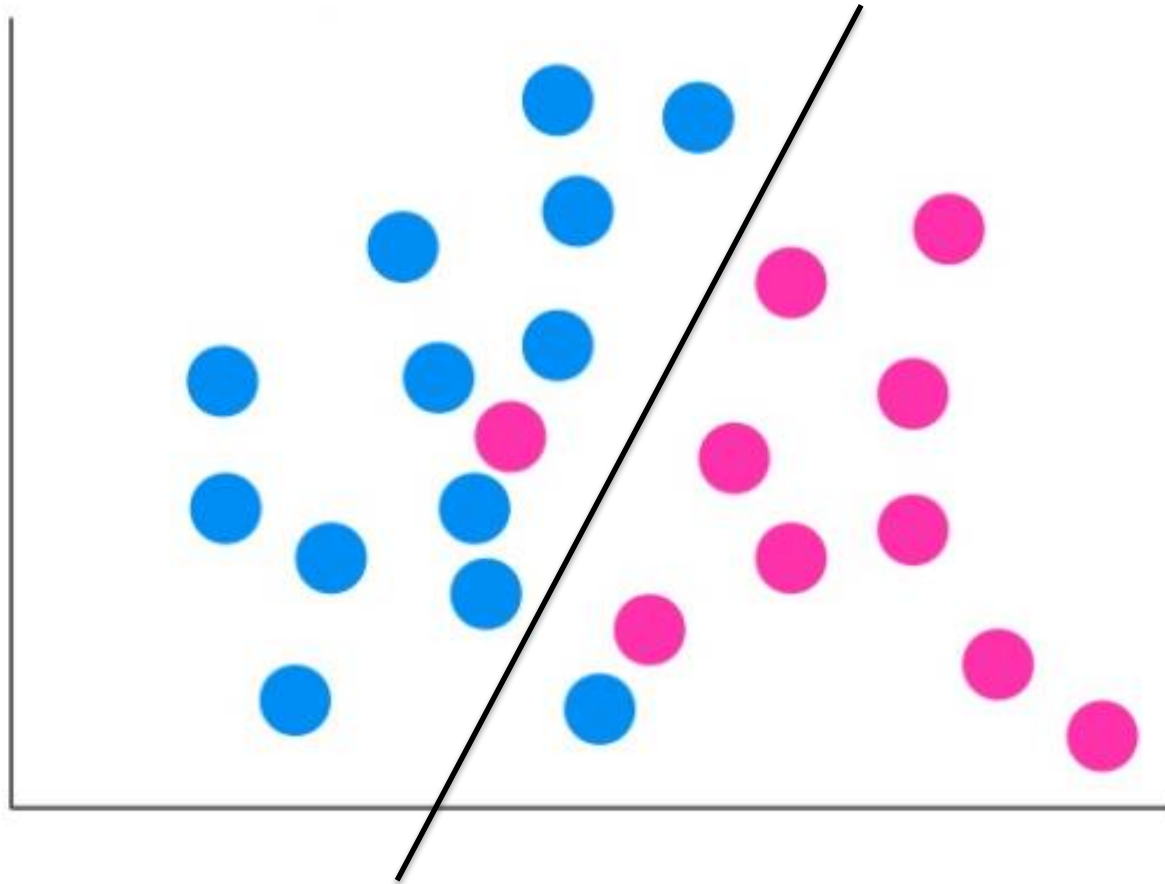
Classification



4 blue dots and 2 red dots are wrongly classified.

But it is better than previous classification.

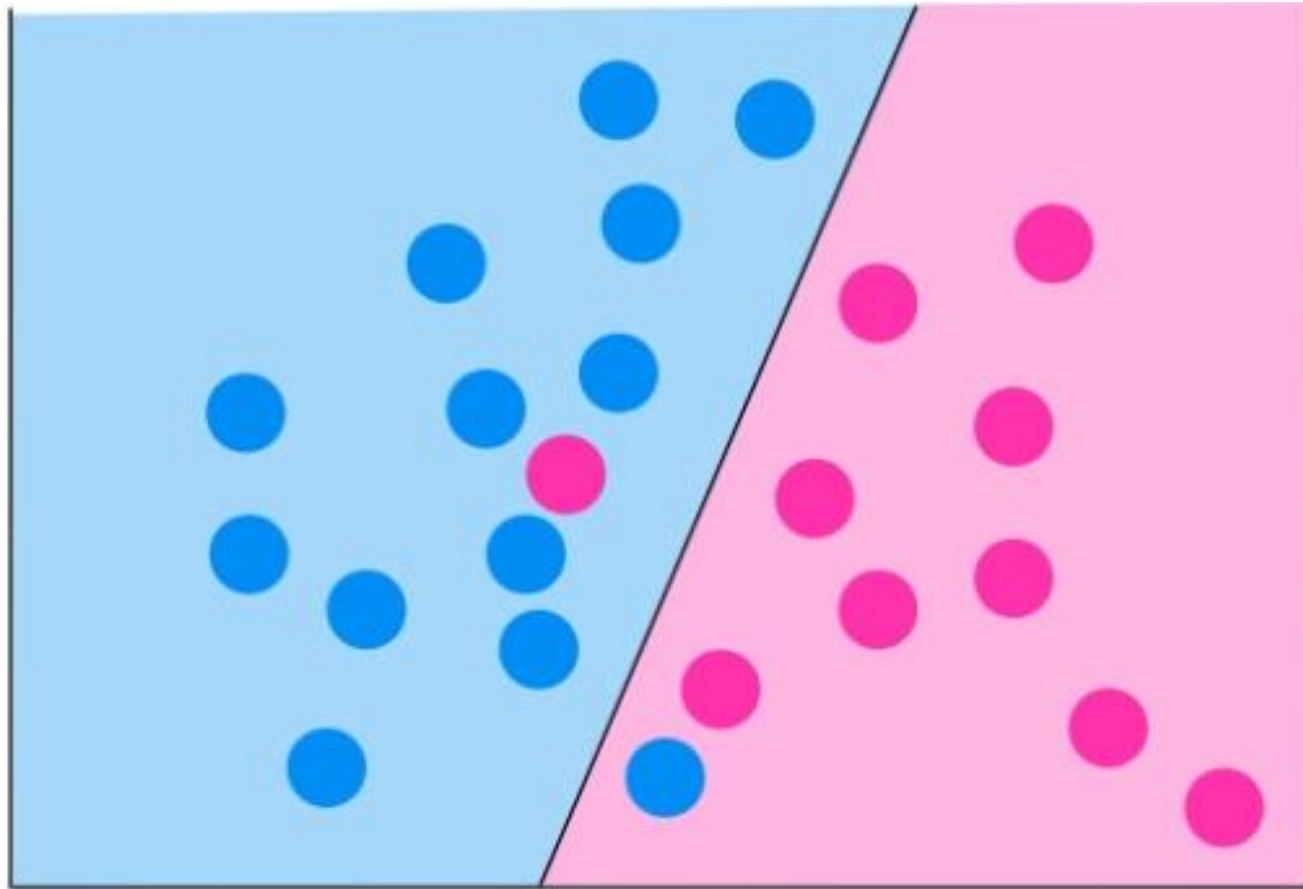
Classification



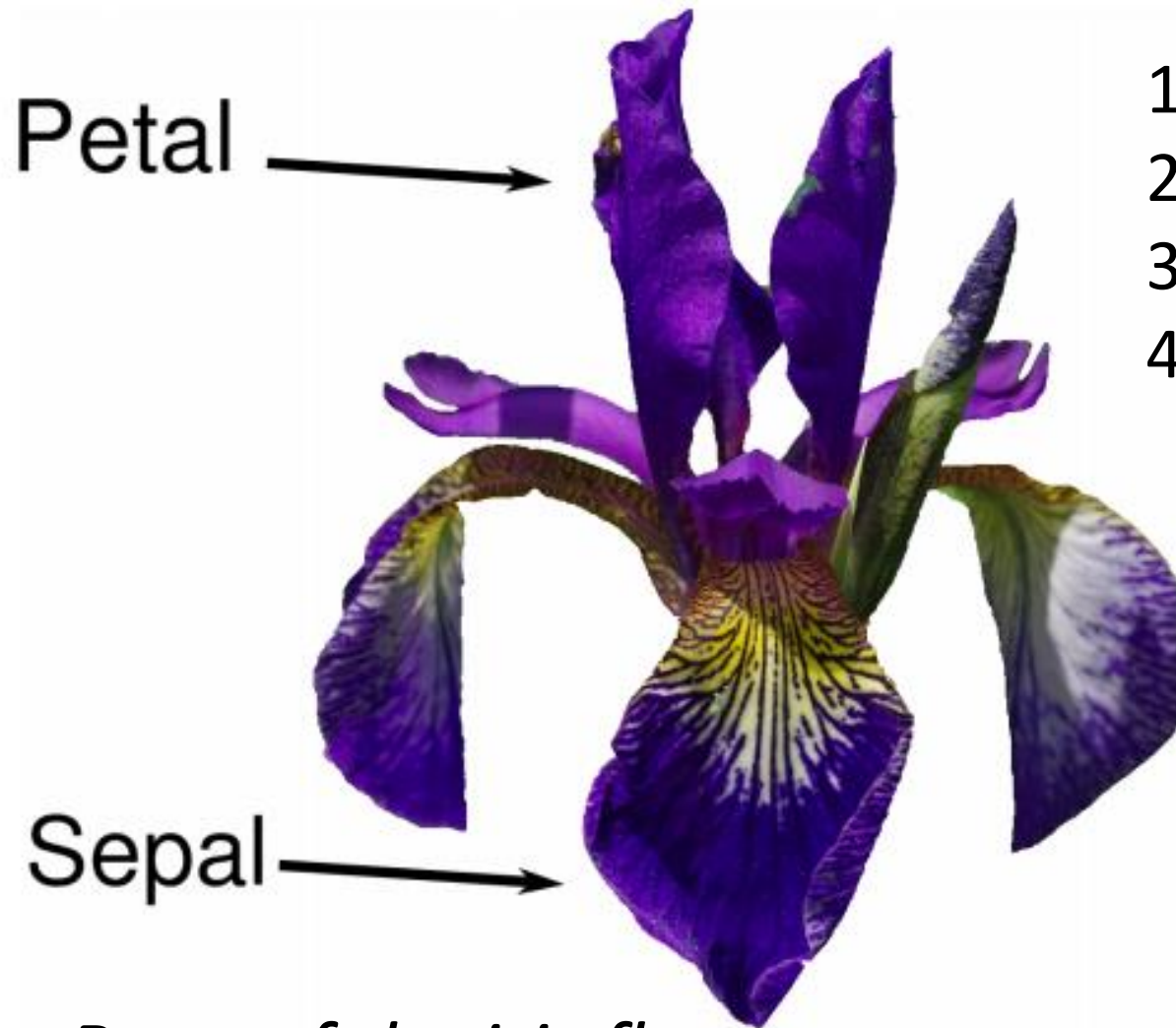
1 blue dot and 1 red dot are wrongly classified.

But it is better than previous classification.

Classification



Classification



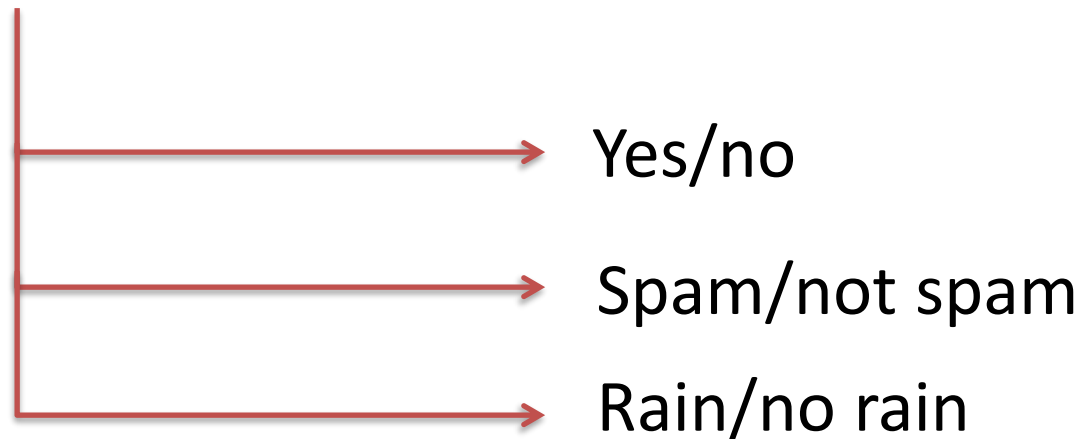
1. length of the petal
2. width of the petals
3. length of the sepal
4. width of the sepals

Parts of the iris flower

Flower species
setosa,
versicolor,
virginica

Classification

- ❑ The goal is to predict a class label (categorical variable)
- ❑ Binary and multiclass classification



Thanks!