

دوره دیتا ساینس کاربردی

Machine Learning

Logistic Regression

× ×
× ×
× ×
× ×

× ×
× ×
× ×
× ×

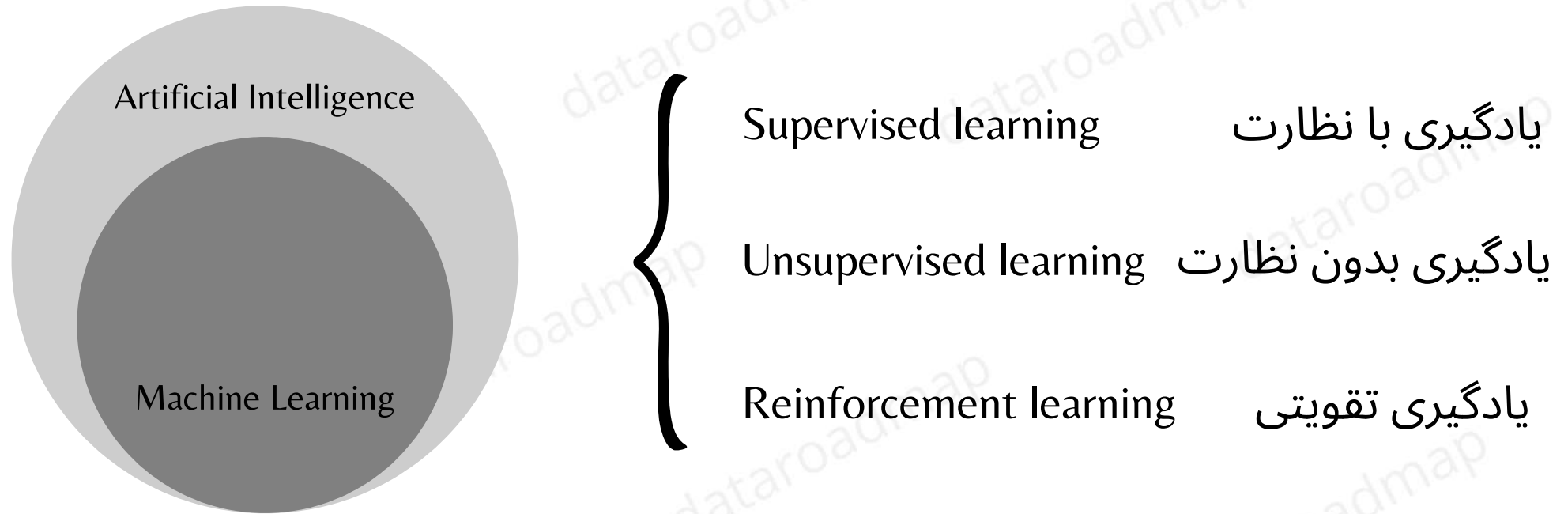
—● dataroadmap ●—

مدرس: مونا حاتمی



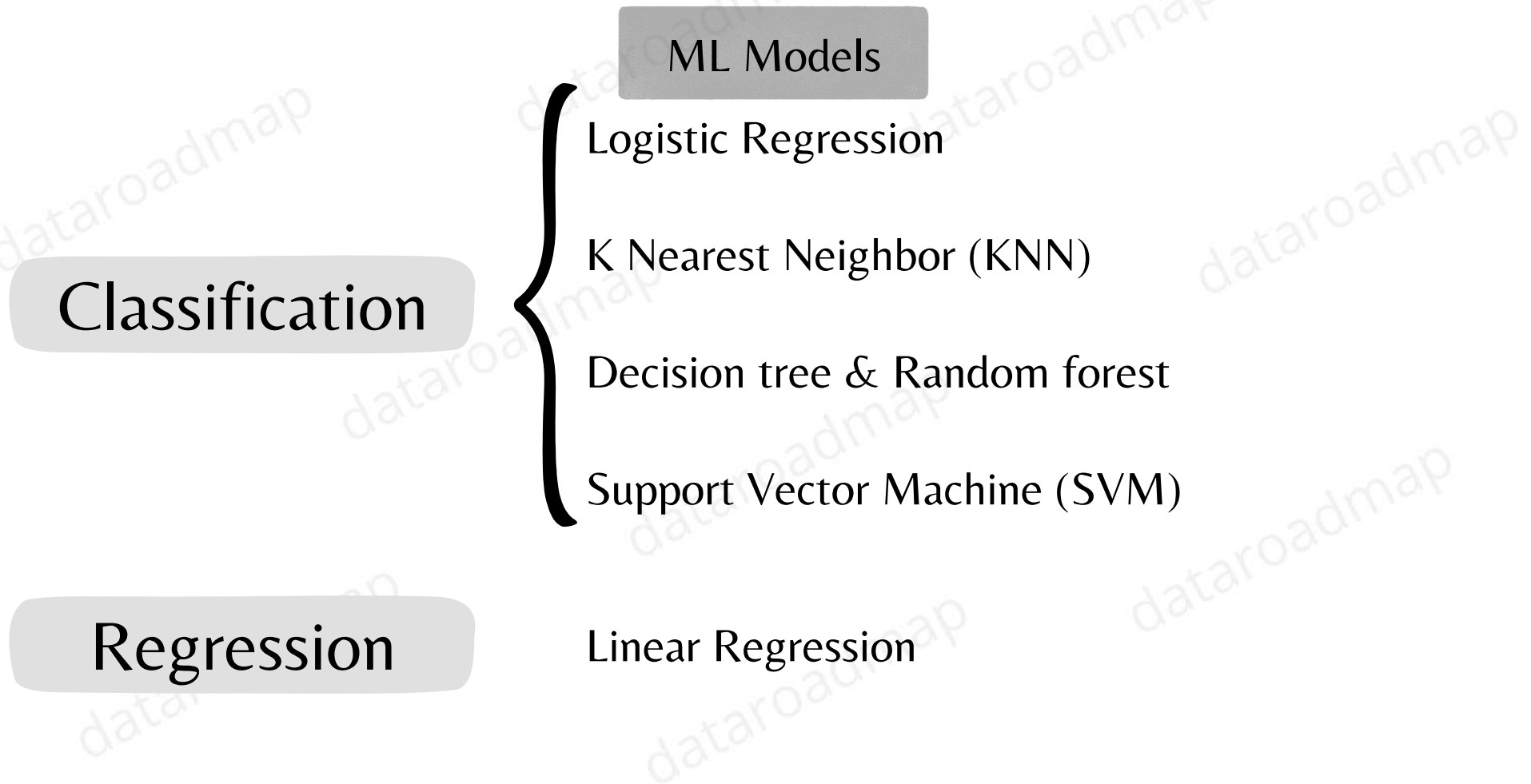
جلسه ششم

What is Machine Learning?

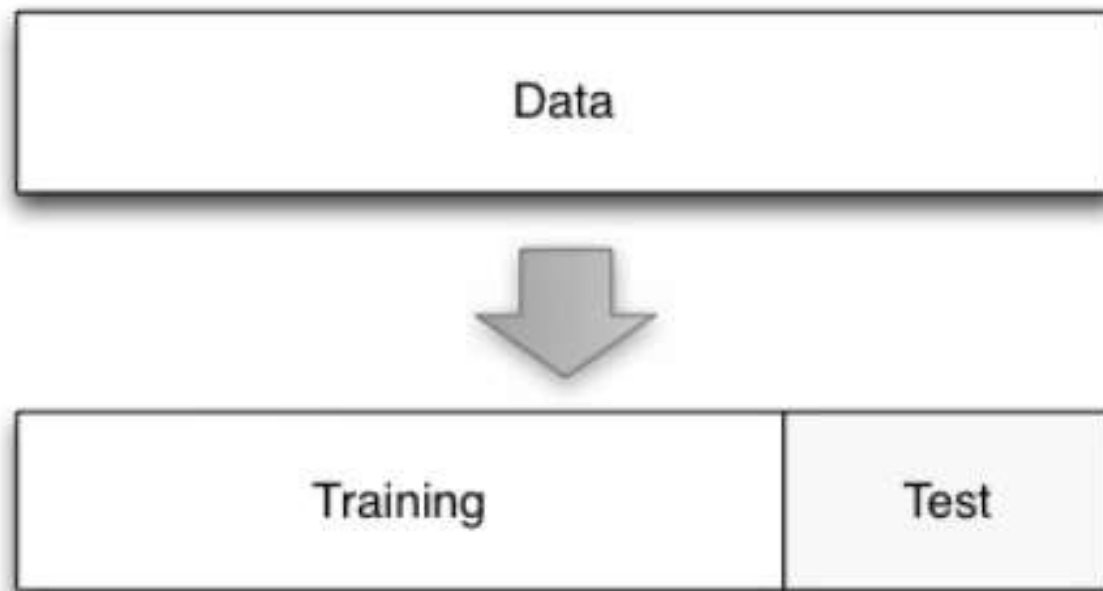


Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output.

Output in Supervised Learning?



Train set- Test set in Supervised Learning



Read the Data

```
import pandas as pd
```

The Data ¶

```
df = pd.read_csv('preprocessed_dataset.csv')
```

```
df.head()
```

[3]:

	Unnamed: 0	PayloadMass	Flights	GridFins	Reused	Legs	Block	ReusedCo
0	0	6104.959412	1	0	0	0	1.0	
1	1	525.000000	1	0	0	0	1.0	
2	2	677.000000	1	0	0	0	1.0	

Non- Null

```
|: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 90 entries, 0 to 89
Data columns (total 89 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Unnamed: 0             90 non-null    int64
1   PayloadMass            90 non-null    float64
2   Flights                90 non-null    int64
3   GridFins               90 non-null    int64
4   Reused                 90 non-null    int64
5   Legs                  90 non-null    int64
6   Block                 90 non-null    float64
7   ReusedCount            90 non-null    int64
8   Class                  90 non-null    int64
9   Orbit_ES-L1           90 non-null    int64
10  Orbit_GEO              90 non-null    int64
11  Orbit_GTO              90 non-null    int64
12  Orbit_HEO              90 non-null    int64
13  Orbit_ISS              90 non-null    int64
```

```
|: df.isnull()
```

```
[22]:
```

	Unnamed: 0	PayloadMass	Flights	GridFins	Reused	Legs	Block	ReusedCount	Class	Orbit_ES-L1	...	Serial_B1048	Serial
0	False	False	False	False	False	False	False	False	False	False	...	False	False
1	False	False	False	False	False	False	False	False	False	False	...	False	False
2	False	False	False	False	False	False	False	False	False	False	...	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False

Non- Null

```
|: ▶ df.isnull().sum()
```

```
[23]: Unnamed: 0      0  
      PayloadMass    0  
      Flights        0  
      GridFins        0  
      Reused          0  
      ..  
      Serial_B1056    0  
      Serial_B1058    0  
      Serial_B1059    0  
      Serial_B1060    0  
      Serial_B1062    0  
      Length: 89, dtype: int64
```

Define X , y

```
X=df.drop('Class',axis=1)  
y=df['Class']
```


Train set- Test set in sklearn library

```
! pip install sklearn
```

```
from sklearn.model_selection import train_test_split
```

[illegible]

Train the model

```
▶ from sklearn.linear_model import LogisticRegression
```

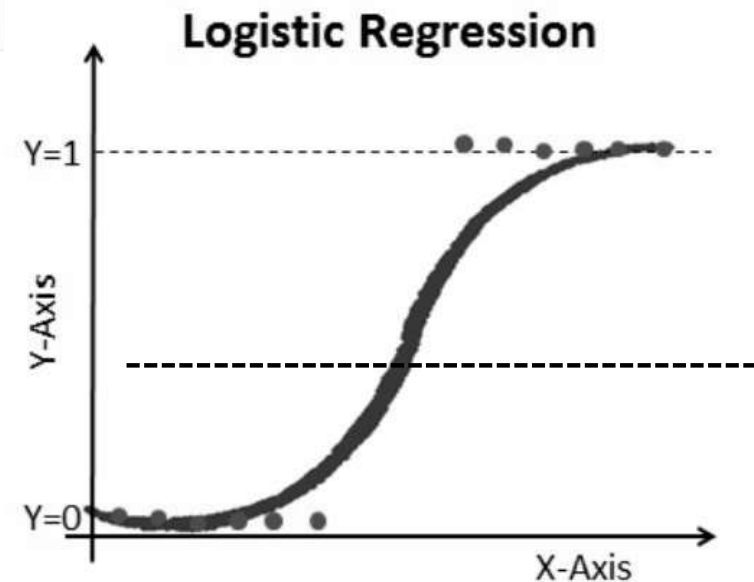
```
▶ logmodel = LogisticRegression()  
logmodel.fit(X_train,y_train)
```

```
] : LogisticRegression()
```

Sigmoid Function

$$p = \frac{1}{1 + e^{-y}}$$

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$



Train the Model

```
logmodel.fit(X_train,y_train)
```

```
LogisticRegression()
```



```
predictions = logmodel.predict(X_test)
```

```
predictions
```

```
: array([0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1], dtype=int64)
```

Prediction

VS

y_test

predictions

```
: array([0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1], dtype=int64)
```

y_test

```
: 50  0
   6  1
   51 0
   54 1
   53 1
   69 1
   32 1
   31 1
   21 1
   88 1
   43 1
   47 0
    3 0
    1 0
   74 0
   16 1
   45 0
   25 1
Name: Class, dtype: int64
```

Evaluation (TP, FP, TN, FN)

```
➤ from sklearn.metrics import confusion_matrix
```

```
➤ confusion_matrix(y_test, predictions)
```

```
|: array([[ 7,  0],  
         [ 1, 10]], dtype=int64)
```

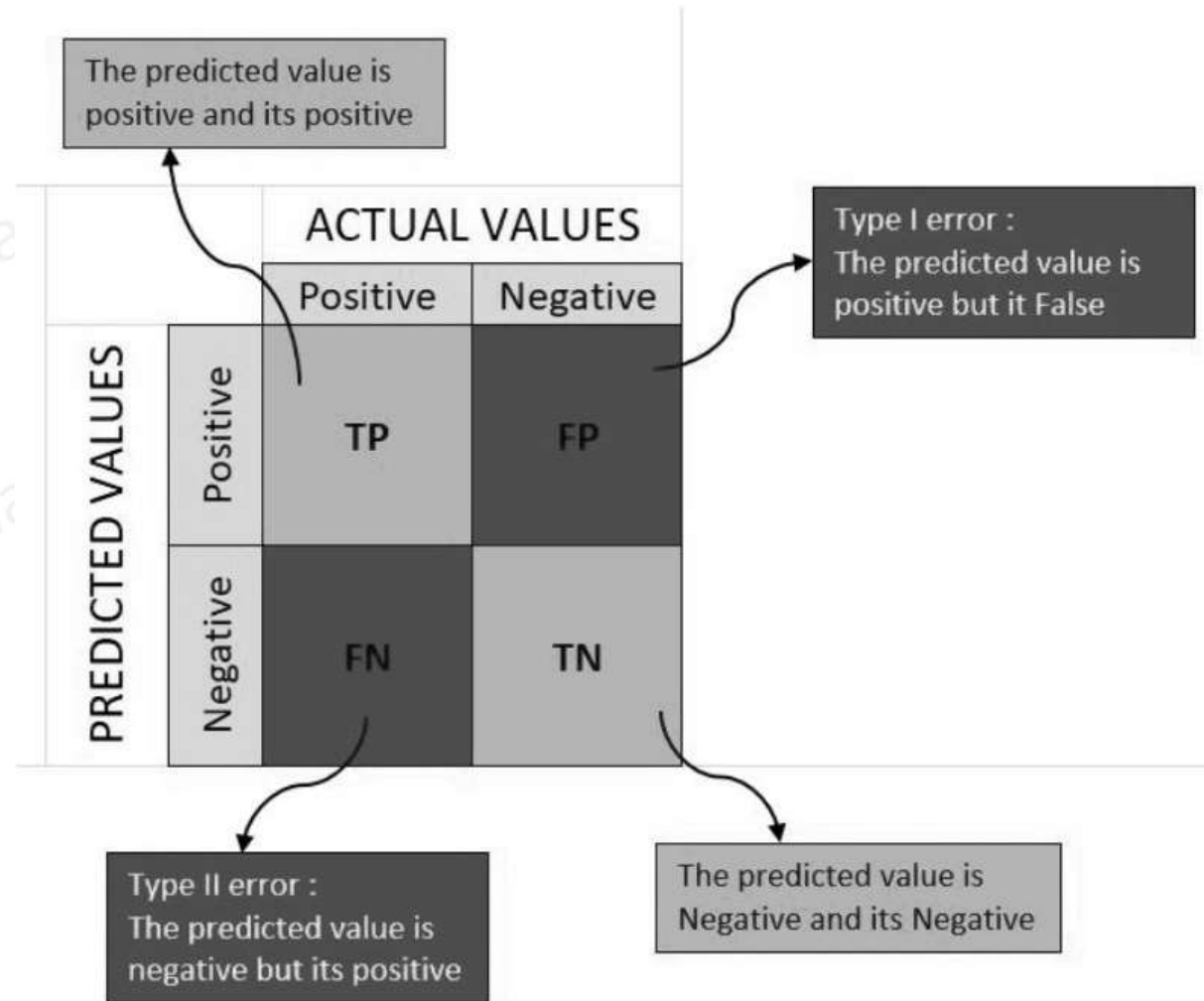
Confusion Matrix

		PREDICTIVE VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	TP	FN
	NEGATIVE (0)	FP	TN

```
array([[ 7,  0],  
       [ 1, 10]], dtype=int64)
```

Confusion Matrix

Error I & Error II



Error I & Error II ????

فرض کنید یک مدل نتیجه آزمایش بیماران را پیش بینی میکند. داشتن بیماری با مقدار 1 (positive) و نداشتن بیماری مقدار 0 (negative) در این صورت کدام نوع خطا بهتر است. چرا؟

Confusion Matrix Error I & Error II

$$\text{Accuracy} = (TP+TN)/(TP+FP+FN+TN) \quad \text{صحت}$$

$$\text{Precision} = TP/(TP+FP) \quad \text{دقت}$$

$$\text{Recall} = TP/(TP+FN) \quad \text{بازخوانی}$$

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

Accuracy Score

```
| from sklearn.metrics import accuracy_score
```

```
| accuracy_score(y_test, predictions, normalize=False)
```

```
17
```

```
| accuracy_score(y_test, predictions, normalize=True)
```

```
0.9444444444444444
```

Classification Report

```
▶ from sklearn.metrics import classification_report
```

```
▶ print(classification_report(y_test, predictions))
```

	precision	recall	f1-score	support
0	0.88	1.00	0.93	7
1	1.00	0.91	0.95	11
accuracy			0.94	18
macro avg	0.94	0.95	0.94	18
weighted avg	0.95	0.94	0.94	18

تمرین: مقادیر جدول 2 را با استفاده از جدول 1 و فرمولهایشان بدست آورید.

PREDICTIVE VALUES

POSITIVE (1) NEGATIVE (0)

POSITIVE (1)	TP	FN
NEGATIVE (0)	FP	TN

1

2

ACTUAL VALUES

```
[[ 7,  0],
 [ 1, 10]],
```

	precision	recall	f1-score	support
0	0.88	1.00	0.93	7
1	1.00	0.91	0.95	11
accuracy			0.94	18
macro avg	0.94	0.95	0.94	18
weighted avg	0.95	0.94	0.94	18

Assignment:

تمرین:

کدهای ارائه شده در درس را در نوتبوک جدیدی انجام داده و در صورت نیاز از نوتبوک هفته ششم استفاده کنید.

برای تمرین بیشتر جلسه نهم دوره منتورینگ دیتاساینس را در کانال یوتیوب مشاهده کنید.

رزومه خود را آپدیت کرده و مهارتهایی که تا به امروز فراگرفته اید را اضافه کنید.