

Recap:

Machine Learning Tree

Supervised learning $x \rightarrow y$

Classification [discrete]
Regression [continuous]

(unlabeled)
unsupervised learning

Clustering
(finding structures/relations in the data).

Reinforcement Learning AI
state action reward

→ Policy based
→ Value based
→ Model based

Iris Dataset

CSV
JSON
excel sheet

labels
features

features				labels
n_1	n_2	n_3	n_4	
Sepal length	Sepal Width	Petal length	Petal Width	species
2.4	3.0	1.8	2.5	setosa

X feature set

Y label set

$y = ['setosa', 'virginica', 'versicolor']$

features				labels
n_1	n_2	n_3	n_4	
5.0	2.9	2.2	1.9	1 setosa
5.1	2.8	2.4	1.6	2 setosa
4.6	3.4	3.1	1.5	3 setosa
5.4	3.7	4.4	1.4	4 setosa
3.2	1.0	0.4	0.1	5 setosa

X-train

Y-train

X-test

Y-test

iris

axis=0
 π_1
 π_2

ML accurate?

!! Break the dataset into two parts.

Train 80%
test 20%

1. Break into X (only features) and y (only labels)

2. Break X and y into training and testing datasets.

accuracy-score(y-test, predictions)

part of actual dataset
made by model

model.
fit
(X_{train} , y_{train})
 $\Rightarrow y_{test} = \text{truth (ground truth)}$
 $\Rightarrow \text{predictions} = \text{made by model}$
 unseen data (X_{test})

In case:
 $y_{test}[i] = \text{'setosa'}$
 and
 $\text{predictions}[i] = \text{'setosa'}$
 we have a correct prediction.

Root Mean Square Error (R.M.S.)

$$\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_{p,i})^2}$$

Classification Accuracy.

(Most simple)

$$\text{Accuracy} = \frac{\text{No. of correct predictions}}{\text{Total no. of predictions made}} \quad [n_{test}]$$

Logarithmic Accuracy

$$\text{Logarithmic Loss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N y_{i,j} * \log(p_{i,j})$$

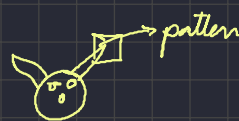
whether sample $i \in j$
 0 or 1

probability

(Stats + Pattern Recognition)

Modern day
ML, CV and
deep learning

Confusion Matrix



Samples: Yes, No

Model Predicted

predictions.

Ground
truth
 y_{test}

$n = \text{sample size}$
 $= 165$

$n = 165$	NO	YES
NO	50	10
YES	5	100

Confusion Matrix

$$\text{Accuracy} = \frac{\text{T.P.} + \text{T.N.}}{N}$$

$$= \frac{100 + 50}{165}$$

$$= \frac{150}{165} = 0.91 \approx 91\%$$

True Positive:

Predicted = YES

Actual = YES

(g.t.) (ground truth)

True Negative:

Predicted: NO

Actual: NO
(gt)

False Positives:

Predicted: YES

Actual (gt): NO

False Negatives:

Predicted: NO

Actual (gt): YES

