

Intro to Al

Radoslav Neychev

Sber, Online 19.07.2021

Outline

- 1. Introduction to Machine Learning, motivation
- 2. ML thesaurus and notation
- 3. Practice
- 4. Q&A



Radoslav Neychev

Assistant professor, ML & DL courses lecturer, MIPT

Visiting lecturer at HSE, Harbour.Space (Barcelona, Spain), Sberbank, Megafon etc.

ex. Yandex-CERN (Research Engineer)

ex. Raiffeisen Bank Russia (Senior Quantitative Analysis Officer)

Motivation, historical overview and

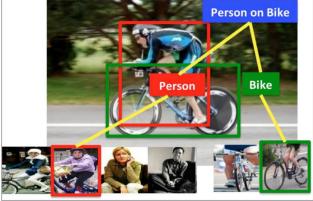
current state of ML and Al

Machine Learning applications



- Object detection
- Action classification
- Image captioning
- ...





Machine Learning applications

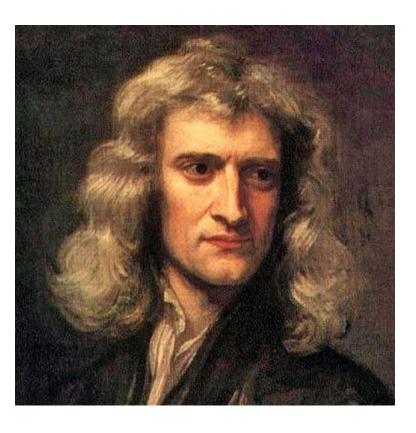


Machine Learning applications



Data — Knowledge

Long before the ML

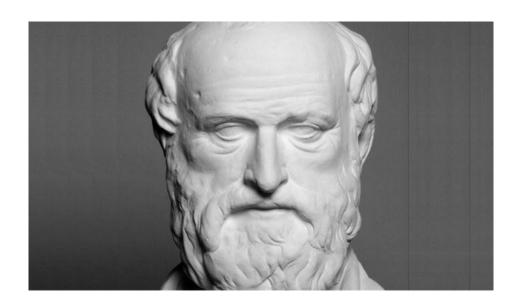


Isaac Newton



Johannes Kepler

Long before the ML



Eratosthenes

FALSE

Denote the dataset.

23

Some

student

Dellote	Denote the dataset.										
		Statistics	Python		Native		Target				
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(passed)				
John	22	2 5	4	Brown	English	5	TRUE				
Aahna	17	4	5	Brown	Hindi	4	TRUE				
Emily	25	5	5	Blue	Chinese	5	TRUE				
Michael	27	3	4	Green	French	5	TRUE				

3 NA

Esperanto

FALSE

Observation (or datum, or data point) is one piece of information.

		Statistics	Python		Native		Target
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(passed)
John	22	5	4	Brown	English	5	TRUE
Aahna	17	4	5	Brown	Hindi	4	TRUE
Emily	25	5	5	Blue	Chinese	5	TRUE
Michael	27	3	4	Green	French	5	TRUE

In many cases the observations are supposed to be *i.i.d.*

independent

Some

• identically distributed

4

5

5

TRUE

TRUE

TRUE

TRUE

FALSE

Fostura (or predictor) represents some special property

reature	or bi	redictor) r	epreser	its some	e speciai p	property.	
		Statistics	Python		Native		T
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(r
John	22	5	4	Brown	English	5	

4

5

3

3

Aahna

Emily

Some

Michael

student

17

25

27

23

	1 - 1 -				[]	<u> </u>	
		Statistics	Python		Native		Target
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(passed)

5 Brown

4 Green

5 Blue

3 NA

Hindi

Chinese

French

Esperanto

These all are features

		Statistics	Python		Native		larget
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(passed)
John	22	5	4	Brown	English	5	TRUE
Aahna	17	4	5	Brown	Hindi	4	TRUE
Emily	25	5	5	Blue	Chinese	5	TRUE
Michael	27	3	4	Green	French	5	TRUE
Some							
student	23	3	3	NA	Esperanto	2	FALSE

FALSE

These all are features

23

Some

student

illese a	These all are realures										
		Statistics	Python		Native		Target				
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(passed)				
John	22	5	. 4	Brown	English	5	TRUE				
Aahna	17	4	5	Brown	Hindi	4	TRUE				
Emily	25	5	5	Blue	Chinese	5	TRUE				
N 4 1 1	0.7					_	TDUE				

ivame	Age	(mark)	(mark)	Eye color	language	rarget (mark)	(passed)
John	22	5	4	Brown	English	5	TRUE
Aahna	17	4	5	Brown	Hindi	4	TRUE
Emily	25	5	5	Blue	Chinese	5	TRUE
Michael	27	3	4	Green	French	5	TRUE

3 NA

3

Esperanto

FALSE

23

3

Some

student

These all are features										
		Statistics	Python		Native		Target			
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(passed)			
John	22	2 5	4	Brown	English	5	TRUE			
Aahna	17	4	5	Brown	Hindi	4	TRUE			
Emily	25	5	5	Blue	Chinese	5	TRUE			
Michael	27	3	4	Green	French	5	TRUE			

3 NA

Esperanto

FALSE

These all are features									
		Statistics	Python		Native		Target		
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(passed)		
John	22	5	4	Brown	English	5	TRUE		
Aahna	17	4	5	Brown	Hindi	4	TRUE		
Emily	25	5	5	Blue	Chinese	5	TRUE		
Michael	27	3	4	Green	French	5	TRUE		

3 NA

Esperanto

Name	Age	(mark)
John	22	
Achno	17	

23

3

Some

student

And even the name is a *feature*

		Statistics	Python		Native		Target
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(passed)
John	22	5	4	Brown	English	5	TRUE
Aahna	17	4	5	Brown	Hindi	4	TRUE
Emily	25	5	5	Blue	Chinese	5	TRUE
Michael	27	3	4	Green	French	5	TRUE
Some							
student	23	3	3	NA	Esperanto	2	FALSE

(despite it might be not informative)

FALSE

The *design matrix* contains all the features and observations. Statistics Dython Mativa Taract

		Statistics	r yulon		INALIVE		iaiyet
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(passed)
John	22	5	4	Brown	English	5	TRUE
Aahna	17	4	5	Brown	Hindi	4	TRUE
Emily	25	5	5	Blue	Chinese	5	TRUE
Michael	27	3	4	Green	French	5	TRUE

Esperanto

3 NA

Some

student

23

3

Features can even be multidimensional, we will discuss it later in this course.

FALSE

Target represents the information we are interested in.

		Statistics	Python		inative		rarget
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(passed)
John	22	5	4	Brown	English	5	TRUE
Aahna	17	4	5	Brown	Hindi	4	TRUE
Emily	25	5	5	Blue	Chinese	5	TRUE
Michael	27	3	4	Green	French	5	TRUE
Some							

3 NA

Esperanto

Target can be either a **number** (real, integer, etc.) – for **regression** problem

23

Target represents the information we are interested in.

		Statistics	Python		Native		Target
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(passed)
John	22	5	4	Brown	English	5	TRUE
Aahna	17	4	5	Brown	Hindi	4	TRUE
Emily	25	5	5	Blue	Chinese	5	TRUE
Michael	27	3	4	Green	French	5	TRUE
Some							
student	23	3	3	NA	Esperanto	2	FALSE

Or a label – for classification problem

Target represents the information we are interested in.

	-						
		Statistics	Python		Native		T
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(
John	22	5	4	Brown	English	5	

5 Blue

3 NA

4 Green

Mark can be treated as a label too (due to finite number of labels:

17 Aahna 5 Brown

5

3

3

25

27

23

1 to 5). We will discuss it later.

Emily

Some

Michael

student

Esperanto

Hindi

Chinese

French

Target (passed)

5

5

TRUE

TRUE

TRUE

TRUE

FALSE

Further we will work with the numerical target (mark)

		Statistics	Python		Native	
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)
John	22	5	4	Brown	English	5
Aahna	17	4	5	Brown	Hindi	4
Emily	25	5	5	Blue	Chinese	5
Michael	27	3	4	Green	French	5
Some student	23	3	3	NA	Esperanto	2

The *prediction* contains values we predicted using some *model*.

		Statistics	Pytnon		Native		Fredicted
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(mark)
John	22	5	4	Brown	English	5	4.5
Aahna	17	4	5	Brown	Hindi	4	4.5

Name	Age	(mark)	(mark)	Eye color	language	larget (mark)	(mark)
John	22	5	4	Brown	English	5	4.5
Aahna	17	4	5	Brown	Hindi	4	4.5
Casily.	25	F	F	Dive	Chinaga	F	5

John	22	5	4	Brown	English	5	4.5
Aahna	17	4	5	Brown	Hindi	4	4.5
Emily	25	5	5	Blue	Chinese	5	5

 $\operatorname{mark}_{ML} = \frac{1}{2} \operatorname{mark}_{Statistics} + \frac{1}{2} \operatorname{mark}_{Python}$

Python marks. So our *model* can be represented as follows:

4

5

4.5

5

3.5

3

The *prediction* contains values we predicted using some *model*. Predicted Statistics Dython Native

		Statistics	ryulon		INALIVE		rredicted
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(mark)
John	22	5	4	Brown	English	5	4.5

4 Green

3 NA

French

Esperanto

Different models can provide different predictions:

3

17

25

27

23

Aahna

Emily

Some

student

Michael

 $\operatorname{mark}_{ML} = \frac{1}{2} \operatorname{mark}_{Statistics} + \frac{1}{2} \operatorname{mark}_{Python}$

5

3

The *prediction* contains values we predicted using some *model*. Pradictad Mativo Ctatiation Dython

		Statistics	Python		ivalive		i redicted
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(mark)
John	22		4	Brown	English	5	1
Aahna	17	4	5	Brown	Hindi	4	5

5 Blue

 $\operatorname{mark}_{ML} = \operatorname{random}(\operatorname{integer from} [1; 5])$

Michael 27 4 Green French Some

Emily

25

Chinese

23 student 3 NA **Esperanto**

Different models can provide different predictions:

5

5

4

3

The *prediction* contains values we predicted using some *model*.

		Statistics	Python		Native		Predicted
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(mark)
John	22	5	4	Brown	English	5	1
Aahna	17	4	5	Brown	Hindi	4	5

5 Blue

4 Green

Chinese

French

Student 23 3 NA Esperanto

Different models can provide different predictions.

5

25

27

Emily

Some

Michael

Usually some **hypothesis** lies beneath the model choice.

Loss function measures the error rate of our model.

Square		Predicted
Square deviation	Target (mark)	(mark)
16	5	1
1	4	5
9	5	2
1	5	4
4	0	2
1	2	3

• **Mean Squared Error** (where y is vector of targets):

$$MSE(\mathbf{y}, \mathbf{\hat{y}}) = \frac{1}{N} ||\mathbf{y} - \mathbf{\hat{y}}||_2^2 = \frac{1}{N} \sum_i (y_i - \hat{y}_i)^2$$

Loss function measures the error rate of our model.

Absolute		Predicted
deviation	Target (mark)	(mark)
4	5	1
1	4	5
3	5	2
1	5	4
1	2	3

• *Mean Absolute Error* (where y is vector of targets):

$$MAE(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N} ||\mathbf{y} - \hat{\mathbf{y}}||_1 = \frac{1}{N} \sum_{i} |y_i - \hat{y}_i|$$

5

5

5

3.5

3

To learn something, our *model* needs some degrees of freedom:

		Statistics	Python		Native		Predicted
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(mark)
John	22	5	4	Brown	English	5	4.5

5 Brown

5 Blue

27 Michael 4 Green French Some 23 3 student

 $\operatorname{mark}_{ML} = w_1 \cdot \operatorname{mark}_{Statistics} + w_2 \cdot \operatorname{mark}_{Python}$

5

Aahna

Emily

17

25

Esperanto 3 NA

Chinese

4.5 Hindi 4

4

5

5

2

4.734

5.101

3.714

3.060

To learn something, our *model* needs some degrees of freedom:

				_			
		Statistics	Python		Native		Predicted
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(mark)
John	22	5	4	Brown	English	5	4.447

5 Brown

5 Blue

Hindi

Chinese

Michael 27 3 4 Green French
Some
student 23 3 NA Esperanto

 $\operatorname{mark}_{ML} = w_1 \cdot \operatorname{mark}_{Statistics} + w_2 \cdot \operatorname{mark}_{Python}$

5

Aahna

Emily

17

25

4

5

5

5

4

3

To learn something, our *model* needs some degrees of freedom:

		Statistics	Python		Native		Predicted
Name	Age	(mark)	(mark)	Eye color	language	Target (mark)	(mark)
John	22	5	4	Brown	English	5	1

5 Brown

Emily 25 5 5 Blue Chinese 27 Michael 4 Green French Some

3

17

23

Aahna

student

$$25$$
 3 4 Green French
 23 3 NA Esperanto
 $23\hat{k}_{ML} = \mathrm{random}(\mathrm{integer\ from\ [1;\ 5]})$

Hindi

Last term we should learn for now is hyperparameter.

Hyperparameter should be fixed before our model starts to work with the data.

We will discuss it later with kNN as an example.

ML thesaurus Recap: Dataset Observation (datum) Feature Design matrix Target Prediction Model Loss function Parameter Hyperparameter

Time for some practice!

Q & A