Introduction to Sequential Data Modelling

Skoltech

Course Structure

Skoltech

Tuesday/Thursday
Lecture - 1/2 hour(s)
Seminar - 1/2 hour(s)

Friday
Test

Instructor
Ekaterina (@kate_nil)

Manager
Alexandra (@alexazdan)

Skoltech

Classical models Block
Focus on Time Series Analysis

22/11 - 06/12

Neural models Block
Focus on Neural Sequential data Modelling

08/12 - 16/12

Projects defence - 20/12

Skoltech

What is the sequential data?

People are the core of everything

Loan application

Alex, 28 years old

Will return a credit?

What data do we need?

How to improve customer experience?

What data do we have?

- Tabular data is hard to collect
- It is prone to errors

A sequence of clients actions - best reflection on what she/he is doing

Alex, 28, man works in retail

Login to a bank app

Watch LOTR

Refuel 1753 r.

Lunch 400 r.

Transfer 2000 r. to Maria

Skoltech

9:34　　　　12:45　　　　18:58　19:00　　　　20:13

Sequential supervised learning problem

Skoltech

The classical supervised learning problem

x - an image
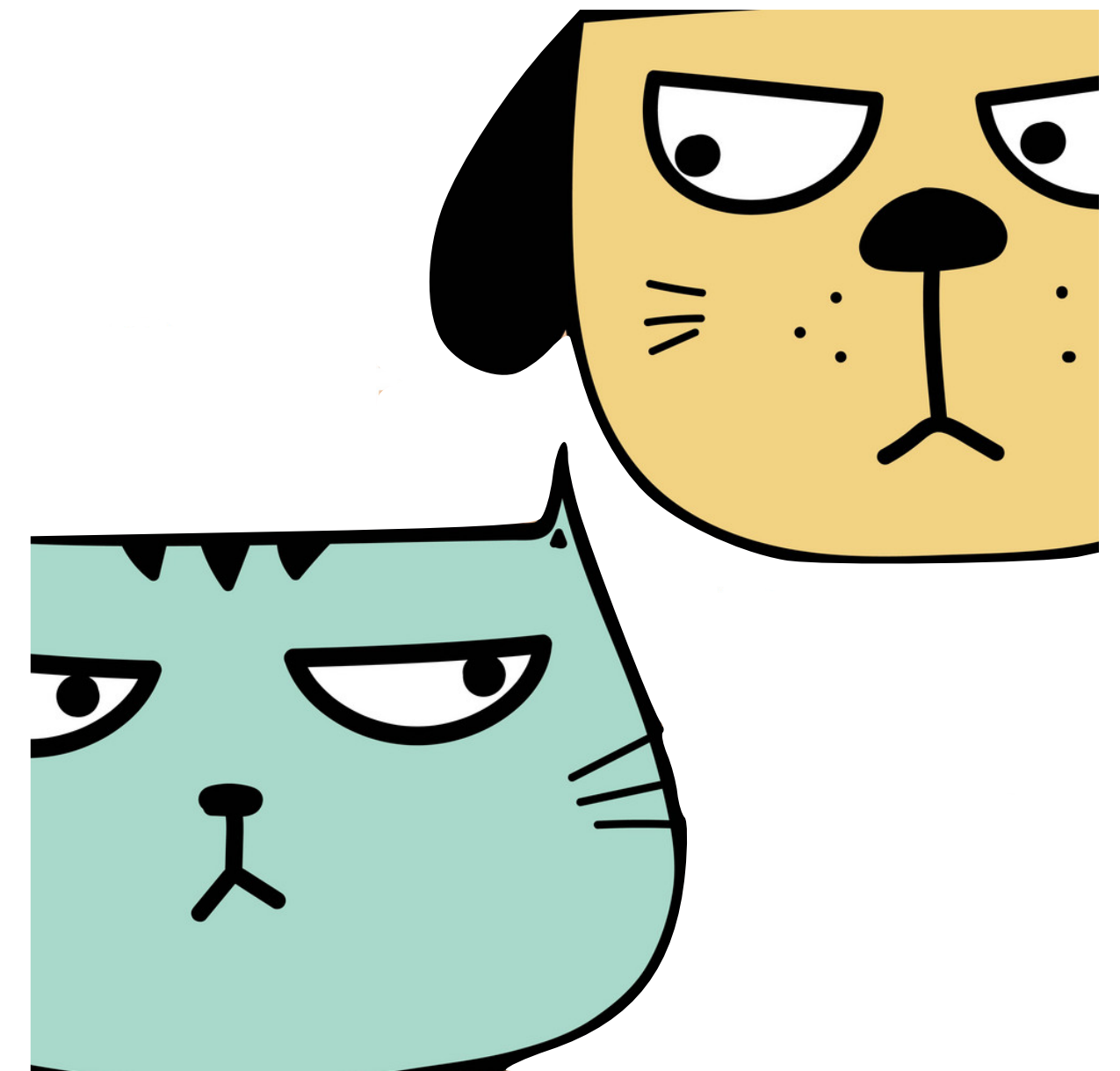
y - a class of an image

$y \in \{0,1\}$

$(x, y) \sim P(x, y)$ (independently and identically)

$\{(x_i, y_i)\}_{i=1}^{N}$ - a set of N training examples

h - classifier, i.e. a function that that maps from images to classes

Learning goal: find an h that correctly predicts the class $y = h(x)$ of new images x

The classical supervised learning problem for sequences

x - a telephone call

y - a class of a call

$y \in \{0,1\}$

x - a part-of-speech

y - a tag of a part-of-speech (noun, verb, adjective, etc.)

$y \in \{0,1,...,N\}$

x - a tweet

y - a class of a tweet

$y \in \{0,1\}$

Not drawn independently and identically (iid) from some joint distribution

Sequences exhibit significant sequential correlation

Skoltech

The sequential supervised learning problem

$\{(x_i, y_i)\}_{i=1}^{N}$ - a set of N training examples

(x_i, y_i) - pair of sequences

$x_i = \langle x_{i,1}, x_{i,2}, \ldots, x_{i,T_i} \rangle$

$y_i = \langle y_{i,1}, y_{i,2}, \ldots, y_{i,T_i} \rangle$

Learning goal: find an h that correctly predicts the class $y = h(x)$ of new sequence x

$x_i = \langle$ you will learn sequence modelling $\rangle$

$yi = \langle$ pronoun verb verb noun noun $\rangle$

Skoltech

Time series models

Skoltech

How to predict the position of a ball at time (T + 1)?

T - 2   T − 1   T

Sequential data is essential for some prediction problems

Skoltech

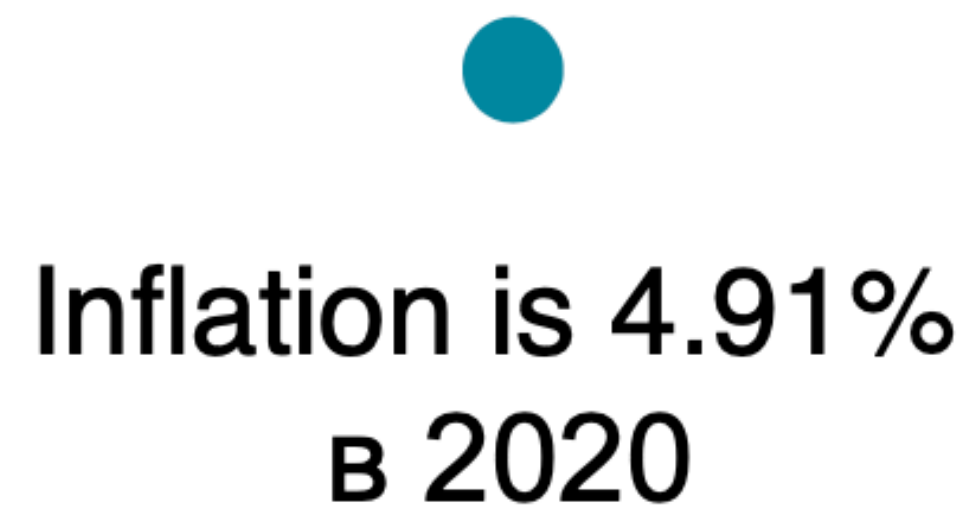# For credit scoring usage of sequential data leads to better models

	Number of input features (↓)	GINI (↑)
Logistic regression	400	0.56
Gradient boosting	7000	0.62
NNs for time series	**12**	**0.66**

Classic methods

Time series models - the right choice

high model quality

little feature engineering

Skoltech

One point-based analytics 😒

Sequential data-based analytics 👍

Inflation is 4.91% в 2020

2018
4.27

Inflation 4.91% in 2020

2.52
2017

3.07
2019

Hard to judge, if everything is good or bad
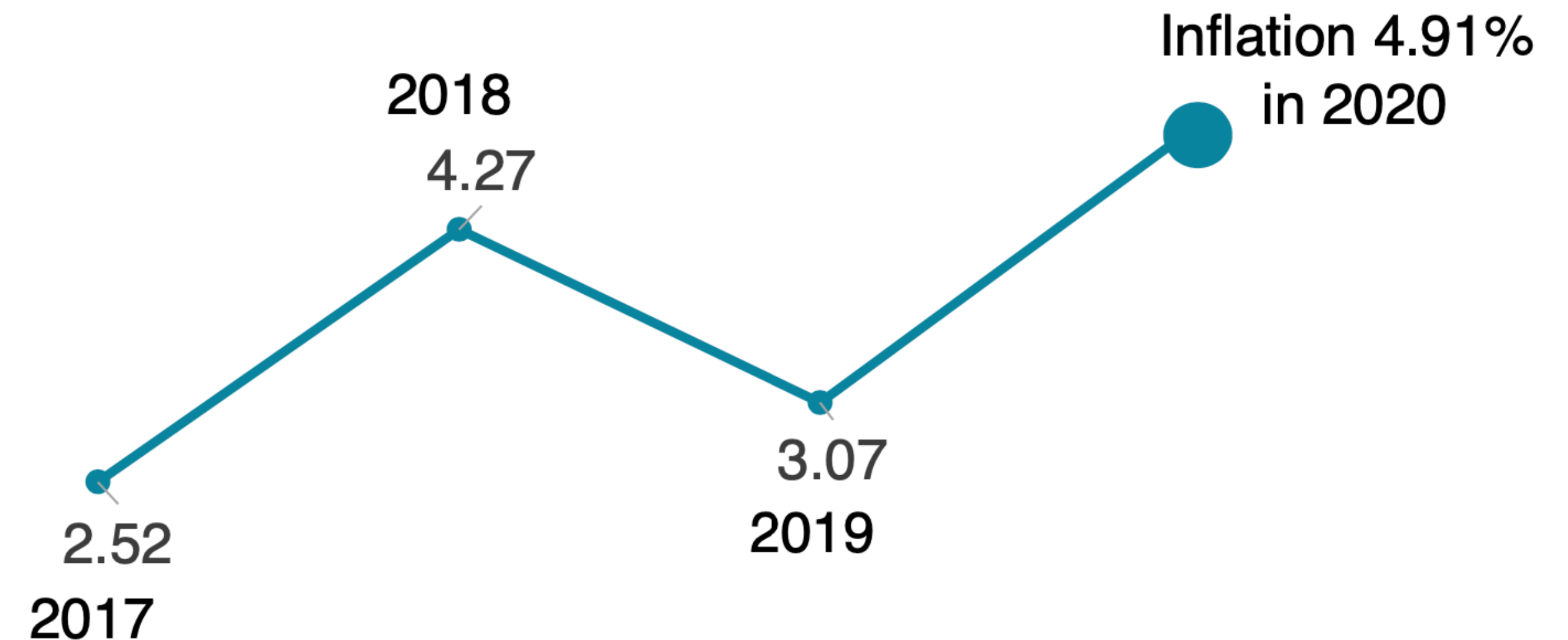
Changes with time is a correct way to look at data and make decisions

Skoltech

Basic Objectives of the Time Series (TS) Modelling

To describe the important features of the time series pattern

To explain how the past affects the future or how two TS can "interact"

To forecast future values of the series

Skoltech

16