

# Machine Learning

## Final Exam

Aleksandr Petiushko

ML Research



# Content

## ① Exam information

# Content

- ① Exam information
- ② Exam problems

# Exam information

- Exam format: **hand-written on paper**
- Exam duration: **1.5 hrs**
- Materials allowed to be used: **hand-written notes (1 sheet of paper)**
- **No one is allowed to use either the phone, laptop or printed materials**
- Exam parts:
  - ① ML Pipeline Design: 60%
  - ② Short questions about the course content: 20%
  - ③ Simple ML-related calculations: 20%

Time for questions



## Exam topics: ML pipeline design

Demonstrate the ability to design the ML pipeline for any given problem. It should consist (but not limited to) of the following sub-steps:

- Clear ML task statement
- Data collection strategy
- Data preparation routines
- Model and loss function design
- Success metrics and eval procedure
- Model selection approach

## Exam topics: ML concepts

Demonstrate the deep knowledge of the following ML concepts:

- Supervised Learning, types of models (high-level)
- Input feature types and dimensionality
- Empirical vs Structural Risk Minimization
- Overfitting vs Underfitting and methods to avoid them
- Cross-Validation
- Model Selection pipeline and why it is important
- Classification vs Regression
- Classification and Regression loss functions
- Classification quality metrics (including accuracy, precision, and recall)
- Regression quality metrics (including MAE, MSE, and RMSE)
- Binary vs Multi-class Classification
- Micro- vs Macro- Averaging for Multi-class Classification
- L1 (Manhattan) and L2 (Euclidian) norms (distances)
- k-NN Classification and k-NN Regression
- Linear Regression: Ridge, LASSO, and Elastic variants

## Exam topics: ML calculations

Additionally, to be able to compute auxiliary things like:

- TP, FP, FN, TN
- TPR, FPR, FNR, TNR
- MAE, MSE, RMSE
- Accuracy, Precision, Recall
- Empirical Risk
- L1 (Manhattan) and L2 (Euclidian) norms (distances), and simple equalities/inequalities based on them



# Exam problems

## ML Pipeline Design

Describe your own task comprising both Classification and Regression problems. State clearly its objective(s). How to collect the training/testing data? How to prepare (pre-process) the collected data? What are input features and outputs of the designed system, their types and ranges? Think of potential ML models to use and loss functions to train. Provide the evaluation procedure and success metrics to select the best model. Suggest the tracking procedure of the possible future issues / ways to overcome them.

## Short questions about the course content

- 1 Definitions of Empirical and Structural Risks. Provide simple examples
- 2 What metrics to use when observing class imbalance? Why? Provide simple examples

## Simple ML-related calculations

- 1 Calculate TP, FP, FN, TN, TPR, FPR, FNR, TNR, Empirical Risk, Precision, Recall if  $GT=(+1,+1,+1,+1,+1,-1,-1,-1)$  and  $prediction=(-1,+1,-1,-1,+1,-1,+1,-1)$
- 2 Weights: (0, 1, 2). Find the regularization parts for LASSO and Ridge Regressions (assuming the multiplicative coefficient is equal to  $\alpha$  and  $\alpha/2$  correspondingly).

Thank you *all*!