Machine Learning
Final Exam

Aleksandr Petiushko

ML Research

Content

1. Exam information

Content

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, both sides can be used)**
- **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%

Exam process

Exam procedure:

- Exam Assignment will be open in **Canvas** for the **time of Exam** (and **only** during this time!)
- You'll need to write Exam Assignment on paper sheets and when you finish, take a photo / convert to pdf and **submit** as a 'File Upload' to Canvas
- You'll have only **1 attempt** to submit, so please be careful!
- After submission, please **deliver** the written sheets to **Professor / Classroom Manager** (to be collected, stored, and probably scanned for better quality)

Remarks on process

- You need to add your name at the top left of the each side of every sheet
- You need to add the page number at the top right of the each side of every sheet (1, 2, 3, 4, 5, ...)
- Allowed: Only one person at a time may leave the room

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

Thank you *all*!