YMLS 2025 fall: Machine Learning Exam Procedure

Exam procedure

- Our exam procedure looks like a small technical interview for an ML Engineer position.
- Exam duration: 30-40 minutes, 4-5 minutes for preparation for each question (you also may start answering right away, if you want).
- You will get 1 question from the first half of the exam set and 1 question from the second half of the questions set.
- Once you know your exam ticket (set of 2 questions) you will have 5 minutes to self-prepare or can start answering right away. The total maximum points you can get for this part is **70 pts.**
- You may be asked beyond your ticket questions for adjacent topics, based on the decision of the examiner. E.g., if your ticket is about Linear Regression, you can get small extra questions on Logistic Regression or Metrics, Naive Bayes.
- During the answer, you need to follow the sequential logic of your narration and give all the formulas, definitions and concepts/theorems you use. The deeper your answer is, the more points you will get.
- You will have paper sheets to write down your answers, please bring your own pen.
- If you want to get 70pts.+ or want to increase your grade for the exam, you may continue with 1 extra question taken randomly from the questions set (except a couple of questions you already answered). You can get extra **30 pts.** for this part. This action may be done only after the examiner's permission – please ask him in advance, and/or he can decide proactively that you need to pass 3rd part on his own. Please, be informed, that you also may <u>decrease your points</u> if you will not be able to properly answer the 3rd question – so please, take this option wisely.
- You will have an exam passing schedule in advance, so you need to be prepared and need to be at the university at least **20 minutes before** your exam timeslot. Otherwise, you may not be allowed to pass the exam.

Complete formula to calculate your final grade for the ML course:

$$\boxed{\mathrm{Final} = \min\left(0.5 \cdot \mathrm{Semester} + 0.5 \cdot \mathrm{Exam}, 100\right)}$$

Important notice: some of the students may get more than 100 for *Semester* or *Exam* components for their excellent results during the semester or exam, as well as might be penalized if the result on the exam will be significantly bad.

Passing criteria: The final grade will still be not greater than 100, refer to the formula above. Passing score for the course is 50 Final pts., at least 3 any homeworks done from 6 (4 HWs and 2 Labs) AND passed the exam with at least 30 pts.

Appeals procedure: Right after the exam you have an option to appeal your results if you are not satisfied with them or have an opinion that the exam process was not fair. You need to report to your Academic Head and describe your appeal case. All the appeal cases will be resolved individually, but please take it wisely as well: your points might be decreased after the appeal procedure.

Worst case scenario: If you will not be able to pass the exam (got less than 30 final pts. during the exam) you will be asked to answer some questions (2-5) from the theoretical minimum questions set (see below at the end of this document). You may get up to 5-10 pts. for each question, but no more than 30pts. total for the exam.

Cheating: If you will be caught cheating or on plagiarism/talking with your groupmates during the exam, you will be deleted right away without the option to appeal or retake the exam.

Retake procedure: There is NO retake procedure provisioned for this subject.

Exam questions list

First part

1. Machine Learning problem statement: Regression, Classification, examples. How to measure quality in classification: accuracy, balanced accuracy, precision, recall, f1-score, ROC-AUC, multiclass extensions.
2. How to measure quality in regression: MSE, MAE, R2. Maximum likelihood estimation, how is it related to regression and classification. Naive bayesian classifier, how does it work
3. K-nearest neighbours classifier, how does it work. Linear regression. Problem statement for the MSE loss function case. Analytical solution. Gauss-Markov theorem. Gradient approach in linear regression.

4. Regularization in linear models: L1 и L2, their properties. Probabilistic interpretation. Logistic regression. Equivalence of MLE approach and logistic loss minimization. Multiclass classification. One-vs-one, one-vs-all, their properties.
5. Support vector machine. Optimization problem for SVM. Kernel trick. Kernel properties.
6. Principal component analysis. Relations to SVD. Eckart-Young theorem. How to apply PCA in practice.
7. Train, validation and test stages of model development.Overfitting problem, ways to detect it. Validation strategies. Cross validation. Data leaks. Bias-variance tradeoff.

Second part

8. Decision tree construction procedure. Information criteria. Entropy criteria, Giny impurity.
9. Ensembling methods. Bootstrap. Bagging. Random Forest, Random subspace method. Boosting and gradient boosting. Main idea, gradient derivation.
10. Matrix calculus and matrix derivatives. How to get the derivative of matrix/dot product. Backpropagation, chain rule.
11. Neural network concept. Fully-Connected layer (FC). Logistic regression as simple NN. XOR problem. Losses for NNs: logistic loss, cross-entropy.
12. Activation functions, their impact on the network, computational complexity. Softmax and LogSoftmax activations, numerical stability. RELU, ELU, LeakyRELU ideas.
13. Optimization methods in Deep Learning. Gradient descent, SGD, it upgrades: Momentum,
RMSProp, Adam. Regularization in Deep Learning: Dropout, Batch Normalization. Differences in training and
evaluation stages.
14. Vanilla Recursive NN cell. Backpropagation through RNN. Vanishing gradient problem. Potential solutions. LSTM/GRU, memory concept, gates ideas.
15. Matrix convolution. Convolutional layer, backpropagation through it. Hyperparameters of Convs. 1x1 convolutions, comparison to FC layers. Max/Average Pooling. Main ideas (high-level) of AlexNet, VGG, Inception (GoogLeNet), ResNet architectures.

Theoretical minimum

You can check yourself before the exam, if you can answer all the questions from the list below, you are in good shape.

If you will not be able to pass the exam (got less than 30 final pts. during the exam) you will be asked to answer some questions (2-5) from this theoretical minimum questions set.
You may get up to 5-10 pts. for each question, but no more than 30pts. total for the exam.

1. Supervised learning problem statement
2. Unsupervised learning problem statement. Provide at least two example problems.
3. What is i.i.d. data?
4. How does a Naive Bayesian Classifier work? Why is it naive?
5. Linear regression model for MSE minimization problem. Write down the formula and the derivative of the loss function w.r.t. weights.
6. Write down a gradient descent step. How to adjust it for huge datasets?
7. What is the likelihood? Where is Maximum Likelihood Estimation (MLE) usually used?
8. What is cross-validation? How does the number of folds affect the validation?
9. What is overfitting and underfitting? How to detect them?
10. What is the difference between parameters and hyperparameters? Provide an example for linear models/decision trees.
11. What is a regularization? What is the difference between L1 and L2 regularization in linear models? Is it the only way to constrain the solution?
12. Does L2 regularization regularize the bias term ($w0$ or b)? Why?
13. Why is it a good idea to normalize data before applying a linear model?
14. Provide a linear classification problem statement. What is a margin?
15. What are precision and recall? How to use them to measure the model quality?
16. Assume the dataset for binary classification is imbalanced, so 95% of data belong to the first class. How to adjust the classification quality measures, to work with such data? Why?
17. What is ROC AUC? How to build the ROC curve?
18. Logistic loss function. How is it related to Maximum likelihood estimation?
19. Support Vector Machine main idea. The optimization functional for linearly separable case.
20. Describe the greedy optimization algorithm for the decision tree.
21. Why can an unconstrained decision tree achieve zero error on the training set with unique objects?
22. How to assign a class label for the object in the tree leaf in classification?
23. How to assign a class label for the object in the tree leaf in regression? Does it depend on the information criterion?
24. What is bagging?25. What is Random Forest? How does it differ from Bagging over decision trees?
26. How are base algorithms being trained in gradient boosting?
27. How does backpropagation work in neural networks? What will be vector by vector derivative?
28. How does the Convolutional layer work? What is the convolution operation?
29. Why fully connected (dense) networks are not the best choice to work with image data? Why do CNNs perform better?
30. How does basic RNN (Vanilla RNN) work?
31. How does dropout work?
32. How do dropout and batch normalization change their behaviour on the inference stage?
33. What is the problem statement for the Principal Component Analysis?