Introduction to machine learning.
Questions for exam.

Innopolis.

April 29 (Friday), 2016.

Each student will be asked 2 random questions from this list with several minutes for preparation. No materials can be used. Bachelors are invited to participate together with masters.

1. Difference between supervised and unsupervised learning. Example of each task.
2. Describe model evaluation with train/test sets, cross validation and leave-one-out techniques.
3. Overfitting and underfitting. Examples of them for classification and regression.
4. Give an example when adding new feature can decrease accuracy on the test set.
5. Give definition of the following models and answer how do the following parameters affect complexity of the model?

model	parameter
K-NN	K
linear regression with regularization	regularizer multiplier λ
decision tree	max depth

6. Give definition of the following models and answer how do the following parameters affect complexity of the model?

model	parameter
decision tree	min samples count in leaf
bagging	number of bootstrap samples
RandomForest	number of sampled features per node

7. How do K-NN and decision tree methods change when they are applied in classification and in regression context?
8. What is curse of dimensionality? Why it appears in K-NN?
9. Give idea how we can optimize K-NN method by removing observations from training set?
10. Give idea how we can optimize K-NN by organizing feature space into nested sequences of rectangles or balls?
11. Explain idea of weighted K-NN. Give examples of weights.
12. Give definition of margin. What is its intuition?
13. Definition of decision tree. How to select splits in each node?
14. Definition of decision tree. When to assign an inner node to leaf node?
15. Definition of decision tree. How to prune redundant nodes?
16. Definition of linear classifier.

17. How to estimate parameters of linear classifier? Write out the idea of misclassification rate approximation and some functions frequently used in this approximation.

18. Give definition of gradient descent and stochastic gradient descent. Motivation for stochastic gradient.

19. Definition of logistic regression using probabilities and using loss function. Show their equivalence.

20. Derive the formula of the distance from point to hyperplane.

21. Regularization - why it is used? Write out L_1 and L_2 regularizers. Which regularization has feature selection capability?

22. How can we control which kinds of features are more/less important in regularization?

23. Describe bagging and RandomForest algorithm. What is their intuition?

24. Which of the following methods always have linear decision boundary separating classes? Which of them under certain conditions may have linear separating boundary?

    (a) K-NN
    (b) Decision tree
    (c) Logistic regression

25. Explain why or why not feature scaling will affect predictions of the following methods:

    (a) K-NN
    (b) Linear regression with regularization
    (c) Decision tree (CART variant that we studied on lectures)
    (d) RandomForest (based also on CART)

26. Gradient descent and stochastic gradient descent. In what cases stochastic gradient descent is better?

27. Why normalization of features is important? Write out possible normalization methods.

28. Ways to represent textual data in vector form using binary indicators, TF and TF-IDF. Intuition of IDF.

29. What is one-hot-encoding and probability encoding? Why it is important to make this transformation?