# CUB Spring 2024. Machine Learning.
# Exam test variant.

Exam rules:

(a) The exam duration time: 120 minutes.

(b) One handwritten A4 page cheat sheet (two-sided) is allowed.

(c) The total number of points is 27. The exam grade is computed as a sum of points for all tasks divided by 22. So for getting 100% it is enough to get in total 22 points out of 27. In order to get the minimum grade 45% it is needed to get in total 10 points.

(d) In tasks with given set of answers the number of correct answers may vary (it could be one answer or multiple answers).

**Task 1 (1 pts)** Choose algorithms that can be used for solving regression problem:

(a) Logistic Regression;

(b) Word2Vec;

(c) Random Forest;

(d) LambdaMART;

(e) Linear Regression.

**Task 2 (2 pts)** Consider the following function:

$$f(\boldsymbol{x}) = \det(\boldsymbol{x}\boldsymbol{y}^T + A).$$

Here $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$. Using the technique of differentials find the gradient of $f(\boldsymbol{x})$.

**Task 3 (1 pts)** Choose the correct statements about Elastic Net regularization:

(a) This regularization is not able to find sparse solutions (zero out weights of some features);

(b) This regularization requires tuning of one regularization coefficient;

(c) This regularization usually shows better performance than $L_1$ regularization in case of presence of highly correlated features;

(d) For linear regression with MSE loss function and this regularization optimal weights can be found analytically.

**Task 4 (1 pts)** Let's consider training a two-class logistic regression without regularization:

$$Q(\boldsymbol{w}) = \frac{1}{N}\sum_{i=1}^{N}\log(1+\exp(-y_i\boldsymbol{w}^T\boldsymbol{x}_i)) \to \min_{\boldsymbol{w}}.$$

Write down one step of stochastic gradient descent algorithm for solving this problem with constant stepsize $\alpha$ and one object in a mini-batch.

**Task 5 (1 pts)** Choose the correct statements about Support Vector Machines:

(a) Weights in decision rule can be computed as a linear combination of points from training set;

(b) Decision rule is not changed in case of excluding some subset of objects from training set;

(c) Training procedure can't be organized with stochastic gradient descent optimizer;

(d) Training procedure can be organized as solution of some linear programming problem.

**Task 6 (2 pts)** In classification problem three objects belong to positive class and 4 objects belong to negative class. A classifier after choosing a threshold on score value makes three errors. What is the maximum possible value of $F_1$-measure?

**Task 7 (2 pts)** Let's consider a two-class classification problem. Suppose that for some dataset a classifier has the following prediction scores

$$[0.7, 0.1, 0.3, -0.2].$$

The corresponding class labels are

$$[1, 1, -1, -1].$$

Compute $F_1$-measure for best threshold.

**Task 8 (1 pts)** Suppose that we build a regression decision tree with variance impurity criterion. Suppose that there are in total $N$ objects in current tree node and $D$ features. What would be computational complexity of exact algorithm for finding optimal split (finding optimal feature number and threshold)?

**Task 9 (2 pts)** Let's consider Decision Tree for classification and Gini index as impurity criterion. This criterion is introduced as follows:

$$H(R) = \min_{c_1,\dots,c_K} \frac{1}{|R|} \sum_{y \in R} \sum_{k=1}^{K} (c_k - [y = k])^2.$$

Here $R$ is a set of objects in current node, $c_j$ is the output value for $j$-th class. Find the optimal value for $c_j$.

**Task 10 (1 pts)** Choose algorithms that build a composition of several models:

(a) Blending;

(b) K-means++;

(c) Bagging;

(d) LightGBM;

(e) t-SNE.

**Task 11 (1 pts)** Choose the correct statements about Gradient Boosting approach:

(a) Gradient Boosting do not overfit with increasing of number of algorithms in composition;

(b) The next algorithm in composition focuses more on objects with large negative value of margin;

(c) The algorithm for composition is usually chosen to have low bias;

(d) For training the next algorithm in composition all objects from the training set are used.

**Task 12 (1 pts)** Let's consider a notion of bias in bias-variance decomposition. Choose methods that allow to reduce bias value:

(a) Increase maximal depth in decision tree;

(b) Decrease maximal depth in decision tree;

(c) Increase regularization coefficient;

(d) Decrease regularization coefficient;

(e) Increase maximal number of decision trees in bagging;

(f) Decrease maximal number of decision trees in bagging.

**Task 13 (1 pts)** Let's consider K-means clusterization algorithm. Suppose that the dataset consists of $N$ objects with $D$ features and we would like to cluster it into $K$ clusters. Find the computational complexity of one iteration of K-means procedure (in case of using standard Euclidean distance).

**Task 14 (1 pts)** Choose the correct statements about K-means clustering algorithm:

(a) The value of minimized criterion is decreasing with increasing of number of clusters;

(b) The algorithm is able to find clusters of arbitrary shape;

(c) The algorithm has incorporated automatic procedure for tuning number of clusters;

(d) The clusterization result depends on clustering initialization.

**Task 15 (1 pts)** Choose algorithms for solving clusterization problem:

(a) t-SNE

(b) DBSCAN

(c) LambdaMART

(d) Random Forest

**Task 16 (2 pts)** Poisson distribution is a discrete probability distribution where a random variable takes values $0, 1, 2, 3, \dots$ with the following probabilities:

$$p(x = k|\lambda) = \exp(-\lambda)\frac{\lambda^k}{k!}.$$

Here $\lambda > 0$ is a parameter of the distribution.

Suppose we have independent samples from this distribution:

$$x_1, x_2, \dots, x_N \sim p(x|\lambda).$$

Find maximal likelihood estimate $\lambda_{ML}$.

**Task 17 (2 pts)** Let's consider a ranking problem with binary relevance. Suppose that relevant objects in dataset have indexes $1, 5, 6, 7, 9$ and ranking algorithm outputs objects with indexes $1, 2, 3, 4, 5$. Compute AveragePrecision@K for all $K = 1, 2, 3, 4, 5$.

**Task 18 (1 pts)** Let's consider ranking problem and $DCG$ criterion, where gain is computed as $2^y - 1$ and discount is computed as $1/i$. Compute normalized $nDCG@5$ if $y_2 = 3$, $y_4 = 4$ and all other $y_i = 0$.

**Task 19 (2 pts)** The entropy for probability distribution is defined as:

$$-\sum_x p(x) \log p(x).$$

Here $p(x)$ is probability for value $x$. For discrete distributions that take values $1, 2, \ldots, N$ find the distribution with maximal entropy, i.e. solve the following optimization problem:

$$
\begin{aligned}
&-\sum_i p_i \log(p_i) \to \max_{p_1,\ldots,p_N},\\
&\sum_{i=1}^{N} p_i = 1,\\
&p_i \geq 0 \ \forall i.
\end{aligned}
$$

**Task 20 (1 pts)** Let's consider two discrete distributions $p$ and $q$ that take values $1, 2, 3$ with the following probabilities:

$$p_1 = p_2 = p_3 = \frac{1}{3}, \ q_1 = \frac{1}{6}, q_2 = \frac{1}{3}, q_3 = \frac{1}{2}.$$

Compute KL divergence between $p$ and $q$.