CUB Spring 2025. Machine Learning.
Mid-Term Exam 28.03.2025

e with printed letters): [redacted]

nber: [redacted]

uration time: 120 minutes.

ritten A4 page cheat sheet (two-sided) is allowed.

umber of points is 27. The exam grade is computed as a sum o
or getting 100% it is enough to get in total 22 points out of 27.
t is needed to get in total 10 points.

th given set of answers the number of correct answers may var
swers).

Give your own examples (not from the lectures) of nominal featu
ures can be colors of picture; red, bleu, green
ures can be ordered list of buying property : rooms,

uppose that you have an overfitted ML model. Choose methods t

Trees increase the maximal depth of a tree; ← incorrect

Trees reduce the maximal depth of a tree;

rease σ parameter in RBF kernel $K(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{x} - \boldsymbol{y}\|^2/(2$

Task 4 (2 pts) Let's consider the following function:

$$f(\boldsymbol{x}) = \operatorname{tr}\left[(A + \boldsymbol{x}\boldsymbol{y}^T)^{-1}\right].$$

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

Task 5 (2 pts) Let's consider the following constrained optimization problem:

$$x^2 - y^2 \to \min_{x,y}$$
$$x^2 + y^2 \leq 1.$$

$F(x) = x^2 - y^2 - \delta(x^2 + y^2 - 1)$

$\rightarrow x^2 + y^2 - 1 = 0$

$\rightarrow x^2 + y^2 - 1 < 0$

Find all stationary points (the points satisfying the KKT theorem).

$\frac{dF(x)}{dx} = 2x - \delta 2x = 0$

$\frac{dF(x)}{dy} = -2y - \delta 2y = 0$

Task 6 (2 pts) Suppose we have independent samples from discrete distribution, where random variable can take values $1, 2, 3$ with the following probabilities:

$$p(x|\alpha): \begin{array}{ccc} 1 & 2 & 3 \\ \alpha & \alpha^2 & 1 - \alpha - \alpha^2 \end{array}$$

Find maximal likelihood estimate α_{ML} if we have in the samples in total 20 ones, 10 twos and 10 threes.

20 10 10

$$\frac{20 \cdot 1 \cdot \alpha + 2 \cdot \alpha^2 \cdot 10 + 10 \cdot 3(1 - \alpha - \alpha^2)}{20 + 20 + 30} = \frac{20\alpha + 20\alpha^2 + 30 - 30\alpha - 30\alpha^2}{70} = \frac{3 - \alpha - \alpha^2}{7}$$

I know it's wrong :(

Task 17 (1 pts) Choose correct statements about Target Encoding procedure:

- ☐ it is some procedure for working with missing/unknown values;
- ☐ in this procedure we need to train some regression model;
- ☐ in contrast to one-hot-encoding this procedure doesn't create additional features;
- ☐ this procedure may lead to leakage of target variable into training set features;
- ☐ this procedure is usually applied for both numerical and discrete features.

Task 18 (2 pts) Let's consider the following optimization problem:

$$\begin{aligned} &\frac{1}{2}x^T Ax \to \min_x, \\ &x_i \leq b \ \forall i \end{aligned}$$

Here $x \in \mathbb{R}^n$, b is some scalar and $A \in \mathbb{R}^{n \times n}$ is some symmetric positively defined matrix. Construct the dual optimization problem.

Task 19 (2 pts) Negative Binomial distribution is a discrete probability distribution where a random variable takes values $0, 1, 2, 3, \ldots$ with the following probabilities:

$$p(x = k \mid q, r) = \binom{k + r - 1}{k} (1 - q)^k q^r.$$

Here $q \in (0, 1)$ and $r > 0$ are parameters of the distribution and $\binom{M}{k}$ – binomial coefficient.

Suppose we have independent samples from this distribution:

$$x_1, x_2, \ldots, x_N \sim p(x|q, r).$$

Find maximal likelihood estimate q_{ML} for fixed r.

(1)

Task 7 (1 pts) Suppose we have two ML models for bank loans approval (scoring models). The first model approves 100 loans, among which 80 were returned and 20 were not returned. The second model approves 50 loans, among which 48 were returned and 2 were not returned. Also, you know that in the dataset there are in total 200 loans that were returned. Which of these two models is better in terms of F_1 measure?

I - precision - $\frac{80}{100} = 0{,}8$ recall - $\frac{80}{200} = 0{,}4$ $F_1 = \frac{2 \cdot 0{,}8 \cdot 0{,}4}{1{,}2} = \frac{1{,}6}{3}$. ← first one is better.

II - precision - $\frac{48}{50} = \frac{96}{100} = 0{,}96$ recall - $\frac{48}{200} = \frac{24}{100} = 0{,}24$ $F_1 = \frac{2 \cdot 0{,}96 \cdot 0{,}24}{1{,}2} = \frac{0{,}96 \cdot 2{,}4}{3} = \frac{1{,}152}{3}$

(0.5)

Task 8 (1 pts) Suppose that for some classifier, the precision is 0.9 and the recall is 0.8. The training set is balanced and has in total 5000 objects. Find false positive rate. ← FPR

precision - 0.9
recall - 0.8
total - 5000

$TP = 0{,}9TP + 0{,}9FP$
$0{,}1TP = 0{,}9FP$
$FP = \frac{TP}{9}$
$TP = 0{,}8TP + 0{,}8FN$
$0{,}2TP = 0{,}8FN$
$TP = 4FN$

$FPR = \frac{FP}{FP+TN} = \frac{FP}{5000 - \frac{45}{4}FP} = \frac{400}{500} = 0{,}8$

$\frac{TP}{FP} = \frac{FN}{TN}$
$TN = FP \cdot \frac{FN}{TP} = \frac{8FP}{4 \cdot 8FP} = \frac{FP}{4}$

$TP + FP + FN + TN = 5000$ $9FP + FP + \frac{9}{4}FP + TN = 5000$ $TN = 5000 - \frac{49}{4}FP$
$\frac{58FP}{4} = 5000$ → $FP = 400$ → $TN = 100$

Task 9 (1 pts) Stochastic gradient descent optimizer with momentum comparing to standard stochastic gradient descent:

(0.62)

- ☐ ~~uses information~~ about second derivatives;
- ☒ makes some kind of averaging of gradients from previous steps; ← correct
- ☒ computes gradient not for the current parameters but for some averaged version of parameters; ← incorrect
- ☐ uses different scaling for every optimizing parameter.

Task 10 (1 pts) Choose correct statements about MAE loss function (mean absolute error) for regression problem:

(0.62)

- ☐ ~~it is a differentiable~~ function;
- ☐ weights of linear regression with this loss function can be found by analytical solution;
- ☒ it allows to find solutions less sensitive to the presence of outliers in dataset; ← correct
- ☒ it is usually faster to compute comparing to standard MSE loss function. ← incorrect

Task 11 (1 pts) Let's consider support vector machine approach without using kernels. Suppose that D – number of features, N – number of training objects, M – number of support vectors. What is computational complexity for making inference on test object?

(0)

Task 12 (2 pts) Suppose that objects in some dataset after sorting them by score value of some two-class classifier have the following class labels (the first object has the highest score):

(2) $[1, 1, -1, -1, 1, -1]$.

$a_1 \geq a_2 \geq a_3 \geq a_4 \geq a_5 \geq a_6$

Find F_1 measure for this classifier for the best threshold.

1) 1 1 -1 -1 1 -1 | -1
TP=3 FP=3 FN=0
Precision = 0,5
recall = 1
$F_1 = \frac{2 \cdot 0.5 \cdot 1}{1+0.5} = \frac{2}{3}$

2) 1 1 -1 -1 1 | -1
TP=3 FP=2 FN=0
Precision = 0,6
recall = 1 $F_1 = \frac{2 \cdot 0.6 \cdot 1}{1.6} = \frac{3}{4}$

3) 1 1 -1 -1 | 1 -1
TP=2 FP=2 FN=1
precision = 0,5
recall = $\frac{2}{3}$ $F_1 = \frac{2 \cdot 0.5 \cdot \frac{2}{3}}{0.5 + \frac{2}{3}} = \frac{4}{7}$

4) 1 1 -1 | -1 1 -1
TP=2 FP=1 FN=1
precision = $\frac{2}{3}$ recall = $\frac{2}{3}$
$F_1 = \frac{2 \cdot \frac{2}{3} \cdot \frac{2}{3}}{\frac{4}{3}} = \frac{2}{3}$

5) 1 1 | -1 -1 1 -1
TP=2 FP=0 ~~recall=1~~ FN=1
precision = 1 $F_1 = \frac{2 \cdot 1 \cdot \frac{2}{3}}{\frac{5}{3}} = \frac{4}{5}$
recall = $\frac{2}{3}$

6) 1 | 1 -1 -1 1 -1
TP=1 FP=0 FN=2
precision = 1 $F_1 = \frac{2 \cdot 1 \cdot \frac{1}{3}}{\frac{4}{3}} = \frac{1}{2}$
recall = $\frac{1}{3}$

Task 13 (1 pts) Let's consider training a two-class linear classification model with exponential loss function $\exp(-M)$, where $M = yw^T x$. Write down one step of stochastic gradient descent algorithm for training weights w with constant stepsize α and one object in a mini-batch.

I'm not gonna solve 13, 80 marspace for 12.

7) | 1 1 -1 -1 1 -1
TP=0 FP=0 FN=3
precision = 1
recall = 0 $F_1 = 0$

(0)

Task 14 (1 pts) Let's consider a two-class classification problem and exponential loss function

(0) $$L(y, z) = \exp(-yz).$$

Here $y \in \{-1, +1\}$. Does this function allow to predict correct class probabilities? Justify you answer.

(0) **Task 15 (1 pts)** For convex loss functions stochastic gradient descent comparing to full batch gradient descent allows:

- ☐ find model parameters with lower value of loss function; ✓ true answer
- ☐ find model parameters faster due to acceleration of one optimization iteration; ✓ incorrect
- ☒ find model parameters faster due to reduction of number of iterations needed for convergence.

Task 16 (1 pts) Choose the correct statements about precision-recall curve:

(0,72)

- ☐ it is monotonic curve w.r.t. increasing of recall value;
- ☐ The area under the curve is more suitable value for the case of unbalanced classes comparing to the area under ROC curve; ← true answer
- ☐ The starting and ending (this depends) point of the curve does not depend on a training set;
- ☒ The area under the curve takes values between 0 and 1. ← correct