# Modern Methods of Decision Makin

Lab 1. Regularized Logistic Regression

Team 3. Semkin, Maksimov, Venediktov

April 2022

## 1 Optimization task

We formulate an optimization task as follows:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{N} f_i(x) + \lambda R(x)$$

where $\lambda > 0$, $f_i(x), R(x)$ - convex

## 2 Formulation for Regularized Logistic Regression

Let $Y = \{-1, 1\}$ - set of classification labels, $N$ - number of objects (observations) Let

$$f_i(x) = \log(1 + \exp(-y_i x^T w_i)))$$

and

$$R(x) = ||x||_2^2$$

Then our optimization task formulates as:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{N} \log(1 + \exp(-y_i x^T w_i))) + \lambda ||x||_2^2$$

where $\lambda > 0$

## 3 Convexity

### 3.1 Regulatization Term

$R(x) = ||x||_2^2 = \sum\limits_{i=1}^{n} x_i^2$ - convex as sum of convex functions

## 3.2 LR Objective

Let $\sigma(z) = \frac{1}{1+\exp(-z)}$. Note that $1 - \sigma(z) = \frac{exp(-z)}{1+\exp(-z)}$

$$\frac{\partial \sigma(z)}{\partial z} = ((1+\exp(-z))^{-1})' = -\frac{1}{(1+\exp(-z))^2}(1+\exp(-z))' =$$

$$= \frac{\exp(-z)}{(1+\exp(-z))^2} = \sigma(z)(1-\sigma(z))$$

Then $\frac{\partial(1-\sigma(z))}{\partial z} = -\sigma(z)(1-\sigma(z))$
Let's use it to find a partial first derivative of $f_i(x)$:

$$\frac{\partial f_i(x)}{\partial x_j} = \frac{1}{1+\exp(-y_i x^T w_i)} \exp(-y_i x^T w_i)(-y_i w_{ij}) =$$

$$= -y_i w_{ij}(1 - \sigma(y_i x^T w_i))$$

Now we find partial second derivative:

$$\frac{\partial^2 f_i(x)}{\partial x_j \partial x_k} = -y_i w_{ij} \frac{\partial(1-\sigma(y_i x^T w_i))}{\partial x_k} = y_i w_{ij} \sigma(y_i x^T w_i)(1-\sigma(y_i x^T w_i)) y_i w_{ik} =$$

$$= y_i^2 w_{ij} w_{ik} \sigma(y_i x^T w_i)(1-\sigma(y_i x^T w_i))$$

Let $H(x)$ - Hessian matrix. As we don't require $j \neq k$, the formula above describes all possible elements of $H(x)$
Now we should check if $H(x) \succeq 0$
By definition, matrix $M$ is semi-definite if $\forall a \neq 0, a \in \mathbb{R}^n$

$$a^T M a \geq 0$$

In our case:

$$a^T H a = \sum_{j=1}^{n} \sum_{k=1}^{n} y_i^2 a_j a_k w_{ij} w_{ik} \sigma(y_i x^T w_i)(1-\sigma(y_i x^T w_i)) =$$

$$= y_i^2 \sigma(y_i x^T w_i)(1-\sigma(y_i x^T w_i)) a^T w_i w_i^T a \geq 0$$

As $y_i^2 = 1, \sigma(y_i x^T w_i) \geq 0, (1-\sigma(y_i x^T w_i)) \geq 0, a^T w_i w_i^T a \geq 0$
Then $f_i(x)$ is convex and $L = \sum_{i=1}^{N} f_i(x)$ is convex as a sum of convex functions