## Theoretical task 1

Solution should be short, mathematically precise and contain proof unless qualitative explanation/intuition is needed. Please underline your final answer for ease of checking. Solution should be handwritten, scanned and sent to mlaticl2019@yandex.ru due Feb 3, 23:59.

- 1. Consider real numbers  $z_1, z_2, ... z_N$ . Find such constant approximation  $\mu$  of these numbers, so that
  - (a)  $\sum_{n=1}^{N} (z_n \mu)^2$  is minimized.
  - (b)  $\sum_{n=1}^{N} |z_n \mu|$  is minimized.

 $\textit{Hint: if a function is convex, zero derivative is a sufficient condition of its global minimum. } \frac{d}{du} \left| u \right| = sign(u).$ 

- 2. Suppose  $x \in \mathbb{R}^D$  is a feature vector. Prove that whitening transformation  $f = \Sigma^{-1/2}(x \mu)$ , where  $\mu = \mathbb{E}x$ ,  $\Sigma = cov[x, x]$ , will give new feature vector f with:
  - (a)  $\mathbb{E}f = \mathbf{0}$  (all zeroes vector)
  - (b) cov[f, f] = I (identity matrix)
- 3. Under what selection of kernel K(u) and bandwidth function h(x) will Nadaraya-Watson regression reduce to K-NN regression?
- 4. Write stochastic gradient descent for linear binary classifier  $\widehat{y}(x) = sign(w^T x)$  with minibatch size=1 for the following losses:
  - (a)  $\mathcal{L}(M) = [-M]_+$
  - (b)  $\mathcal{L}(M) = \ln(1 + e^{-M})$

 $M = w^T xy$  is the margin.

5. Consider finding PCA components from a sequence of optimization tasks, applied to design matrix  $X \in \mathbb{R}^{NxD}$ . You know that

$$\begin{cases} \|Xa_1\|^2 \to \max_{a_k} \\ \|a_1\| = 1 \end{cases}$$

gives eigenvector, corresponding to largest eigenvalue. Prove that

$$\begin{cases} \|Xa_k\|^2 \to \max_{a_k} \\ \|a_k\| = 1 \\ a_k^T a_1 = \dots = a_k^T a_{k-1} = 0 \end{cases}$$

will give eigenvector of  $X^TX$  corresponding to k-th largest eigenvalue.

6. Derive analytical solution for weighted regression:

$$\sum_{n=1}^{N} w_n \left( x_n^T \beta - y_n \right)^2 \to \min_{\beta \in \mathbb{R}^D}$$

in terms of matrix of weights diagonal matrix  $W = diag\{w_1,...w_N\} \in \mathbb{R}^{NxN}$ , design matrix  $X \in \mathbb{R}^{NxD}$  and outputs vector  $Y \in \mathbb{R}^{Nx1}$ , where D is the number of features.

1