

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 mlati2019@yandex.ru due Feb 3, 23:59.

1. Consider real numbers z_1, z_2, \dots, z_N . Find such constant approximation μ 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.

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

2. Suppose $x \in \mathbb{R}^D$ is a feature vector. Prove that whitening transformation $f = \Sigma^{-1/2}(x - \mu)$, where $\mu = \mathbb{E}x$, $\Sigma = \text{cov}[x, x]$, will give new feature vector f with:

(a) $\mathbb{E}f = \mathbf{0}$ (all zeroes vector)

(b) $\text{cov}[f, f] = I$ (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 algorithm for binary linear classifier $\hat{y}(x) = \text{sign}(w^T x)$ with minibatch size=1 for the following losses:

(a) $\mathcal{L}(M) = \max\{-M, 0\}$

(b) $\mathcal{L}(M) = \ln(1 + e^{-M})$

where $M = w^T x y$ is the margin.

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

$$\begin{cases} \|X a_1\|^2 \rightarrow \max_{a_k} \\ \|a_1\| = 1 \end{cases}$$

gives eigenvector, corresponding to largest eigenvalue. Prove that

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

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

6. Derive analytical solution for weighted regression:

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

in terms of matrix of weights diagonal matrix $W = \text{diag}\{w_1, \dots, w_N\} \in \mathbb{R}^{N \times N}$, design matrix $X \in \mathbb{R}^{N \times D}$ and outputs vector $Y \in \mathbb{R}^{N \times 1}$, where D is the number of features.