Theoretical task 2

All solutions should be short, mathematically precise and contain proof unless qualitative explanation/intuition is needed.

1. Consider real numbers $z_1, z_2, ...z_N$. 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. $\frac{d}{du}|u| = sign(u)$.

2. Consider fitting base model m of gradient boosting with loss $\ln(1 + e^{-F(x_n)y_n}), y \in \{+1, -1\}$. Write out the training set to which the base model will be fitted. How can you interpret the targets?

   Hint: does the target depend on margin?

3. Consider a layer of neurons $x \in \mathbb{R}^N$ in multilayer perceptron matched to the next layer using $y = w_0 + W \times ReLU(x)$, $y, w_0 \in \mathbb{R}^M, W \in \mathbb{R}^{MxN}$. Assume bias vector w_0 consists of equal values and weight matrix W consists of equal values. Find these values for w_0 and W to make each neuron from y to have zero mean and unit variance. Assume each neuron x_i takes random values -1 with probability 1/2 and -1 with probability 1/2 and is independent from values of other neurons from x.

   Hint: use central limit theorem.

4. Consider 2 subsequent layers in the middle of multi-layer perceptron [optimized with loss function $\mathcal{L}(\widehat{y}, y)$]: 1st layer gets inputs $\left\{I_i^1 = \sum_i w_{ij} O_i^0\right\}_i$ from outputs of previous layer $\left\{O_i^0\right\}_i$. Outputs $\left\{O_i^1\right\}_i, O_i = f(I_i^1)$ are then fed into 2nd layer inputs $\left\{I_k^2 = \sum v_{jk} O_j\right\}_k$ and produces outputs $\left\{O_k^2\right\}_k$, $O_k^2 = f(I_k^2)$. Suppose weights w_{ij} and v_{jk} for all k, j are initialized to some common constant value c. Prove that weights

   (a) $v_{jk} = const(j)$ after each step of stochastic gradient descent.

   (b) $w_{ij} = const(j)$ after each step of stochastic gradient descent.

   ($const(u)$ means that value remains constant as u get changed)

   Hint: note that $\mathcal{L}(v_{jk}) = \mathcal{L}(O_k^2(I_k^2(v_{jk})))$, $\mathcal{L}(w_{ij}) = \mathcal{L}(\left\{O_k^2(I_k^2(O_j^1(I_j^1(w_{ij}))))\right\}_k)$

5. Consider GAN with true data distribution $p(x)$ and generated data distribution $q(x)$, so log-likelihood is $V(G, D) = \int_x p(x) \log(D(x)) dx + q(x) \log(1 - D(x)) dx$. Let generator and corresponding generator distribution be fixed. Find functional form of optimal discriminator $D(x)$.

   Hint: consider log-likelihood maximization independently for every object x.