Theoretical task 2

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 8, 23:59.

1. Prove that function $K(x, x') = e^{-\gamma\langle x-x', x-x'\rangle}$, $\gamma > 0$ is a valid Mercer kernel.

2. Draw a neural network (structure, weights, thresholds), implementing a XOR function for binary inputs, shown below:

x^1	x^2	x^1 XOR x^2
0	0	0
0	1	1
1	0	1
1	1	0

   The network is supposed to use only $\mathbb{I}[u \geq threshold]$ activation functions.

3. Convex criteria are good, because 1) local minimum is always a global minimum 2) zero gradient is sufficient for that point to be minimum. Is SVM criterion (consider unconstrained maximization problem for linear SVM) a convex problem (in terms of all optimized parameters jointly)? Why?

   Hints:

   - definition: functon is convex if it satisfies $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2)$ for $\forall x_1, x_2 \in domain(f), \forall\alpha \in [0, 1]$.
   - is convex function depending on linear function convex?
   - is sum of convex functions convex?

4. Suppose that you have a binary random classifier, assigning probabilities

$$\begin{aligned} p(y = +1|x) &= \xi \\ p(y = -1|x) &= 1 - \xi \end{aligned}$$

   where ξ is a random variable uniformly distributed on $[0, 1]$ independent of x.

   (a) Suppose you assign class when $p(y = +1|x) \geq \mu$ for some threshold μ. What will be $TPR(\mu)$ and $FPR(\mu)$?

   (b) Plot the ROC curve for this classifier.

5. Consider binary classification performed with M classifiers $f_1(x), ...f_M(x)$. Let probability of mistake be constant $p \in (0, \frac{1}{2})$: $p(f_m(x) = y) = p\,\forall m$ and suppose all models make mistakes or correct guesses independently of each other. Let $F(x)$ be majority voting combiner. Prove that $\forall(x, y)\, p(F(x) \neq y) \to 0$ as $m \to \infty$.

   Hint: use central limit theorem. Consider random variable $\eta = \frac{1}{M}\sum_{m=1}^{M} \mathbb{I}[f_m(x) \neq y]$. How ensemble error and value of η are related?

6. Write out the targets z_n for gradient boosting for each of the following losses:

   (a) $e^{-F(x_n)y_n}$, $y \in \{+1, -1\}$.

   (b) $\ln(1 + e^{-F(x_n)y_n})$, $y \in \{+1, -1\}$.