Theoretical task 3

due 9:00 February 5 (Friday).

Remark: No late submissions allowed this time. All solutions should be short, precise and clearly written.

1. Prove that if particular feature x^i has arbitrary continuous distribution with cumulative distribution function $F(u)$, then monotonous transformation with F will yield uniformly distributed feature:

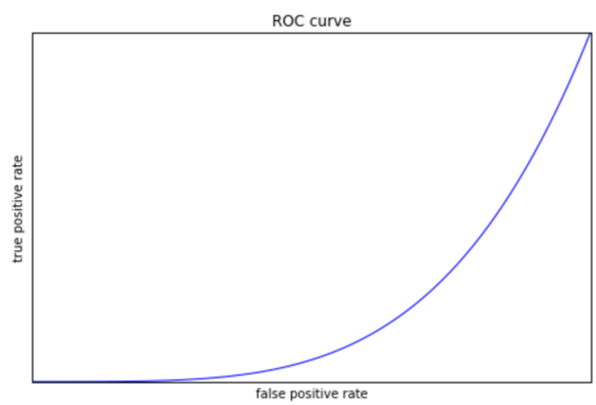
$$F(x^i) \sim Uniform[0,1]$$

2. Suppose that you have a 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. Plot the ROC curve for this classifier and justify your result.

3. Suppose that you have a classifier with the convex ROC curve lying below the line $y=x$ and shown here:



   How can you make this classifier yield you a higher AUC than the random classifier from task (2)? Justify your solution.

4. Suppose your training set consists of N samples and you generate bootstrap pseudosample of the same size.
   (a) What is the probability, that a particular observation will not appear in the bootstrap pseudosample at all?
   (b) What is the limit of this probability as $N \to \infty$?

5. Under what selection of $h(x)$ and $K(u)$ will Nadaraya-Watson regression transform to K-nearest neighbours regression?

6. Explain, why the number of SVM misclassifications, obtained from leave-one-out validation is no greater than the number of support vectors?

7. Prove that polynomial kernel $K(x, x') = (\alpha\langle x, x'\rangle + \beta)^M$ and Gaussian kernel $K(x, x') = e^{-\gamma\langle x-x', x-x'\rangle}$ $(\alpha > 0, \beta > 0, \gamma > 0,\ M = 1, 2, 3, ...)$ are valid Mercer kernels.

8. Draw a neural network (structure, weights, thresolds), 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 $\mathbb{I}[u \geq \mathit{threshold}]$ activation functions.