Theoretical task 4: SVM, neural networks, compositions.

Due Wednesday, January 28.

1. Prove that leave-one-out error rate for SVM E_{LOO} will always be less than or equal to the number of support vectors of SVM trained on the full sample, divided by the size of the full sample N:

$$E_{LOO} \leq \frac{\#SV}{N}$$

2. Suppose your algorithm gives high error rate both on training and on test set.
   (a) Does this situation correspond to high bias (underfitting) or high variance (overfitting)?
   (b) What of the following actions will help?
      i. add regularization, if not present
      ii. increase regularization coefficient leveraging regularization
      iii. add new features
      iv. add polynomial transformations of original features
      v. if algorithm is SVM, increase C parameter
      vi. if algorithm is SVM with RBF kernel, increase γ parameter
      vii. if algorithm is SVM with polynomial kernel, increase polynomial power d
      viii. if algorithm is gradient boosting, decrease the number of steps M
3. Give explanation why original optimization task for SVM with possible misclassifications (penalized by ξ_i) can be equivalently rewritten as regularized misclassification cost minimization of the form:

$$\sum_{i=1}^{N} \mathcal{L}(M(x_i, y_i|w, w_0)) + \frac{1}{2C}||w||^2 \to \min_{w, w_0, \xi}$$

where $||x||$ is euclidean norm, $\mathcal{L}(M)$ is error approximation function equal to $\mathcal{L}(M) = \max\{0, 1-M\}$ and $M(x_i, y_i|w, w_0) = y_i(w^T x_i + w_0)$ is the margin.
4. Write out the form of neural network with exact weights, which has inputs x_1, x_2 (together with constant input on each layer of the network) and which outputs the following function: $\mathbb{I}[-1 \leq x_1 \leq 1] \times \mathbb{I}[-1 \leq x_2 \leq 1]$, where $\mathbb{I}[\cdot]$ is the indicator function. The specified network should have 2 layers (1 hidden+1 output layer) and use $g(x) = \mathbb{I}[x \geq 0]$ activation function.