Machine Learning 2 - HW3

Submission: by 26/01/23 23:55.

1. Generative Models

   You are given the MNIST dataset for training. Let's consider a case in which it turns out there are very few 'bold' 1's in the training set. Your task is to build a model that given a regular image of a 1, generates several 'bold' versions of it.

   a. Describe your model and its architecture.
   b. Describe training details (loss and training scheme)
   c. Describe the generation method to fulfill the task.
   d. What are your assumptions?
   e. What are the limitations of your model, if any?

   Note: you do not need to code your answer, nor submit any experiment supporting it.

2. Rademacher Complexity

   Let $F \circ S$ be the set of all possible evaluations a function $f \in F$ can achieve on a sample S.

   Then, the Rademacher Complexity of F w.r.t. S is

   $$R(F \circ S) = \frac{1}{m} E_{\sigma \sim \{\pm 1\}^m} [sup_{f \in F} \sum_{i=1}^{m} \sigma_i f(z_i)] ,$$

   where the expectation is taken over m i.i.d Rademacher random variables that follow the Rademarcher distribution, i.e.,$P(\sigma_i = 1) = P(\sigma_i = -1) = 1/2$.

   a. Prove that for $\hat{F} = c \cdot f$, for $f \in F$ and a scaler $c \in R$ it holds that

   $$R(\hat{F} \circ S) = |c| R(F \circ S).$$

   Hint: consider $c \geq 0$ and $c < 0$ and try to generalize the expression.

b. Prove that for $\hat{F} = c \cdot f + b$, for $f \in F$ and, a scaler $c \in R$ and a vector $b \in R^m$ it holds that

$$R(\hat{F} \circ S) = |c|\, R(F \circ S).$$