

## Problem Set 2

Student: Brando Miranda

**Problem 1** Please write your analysis on Problem 1 here

**Problem 2** Please write your analysis on Problem 2 here

**Problem 3** Please write your analysis on Problem 3 here

**Problem 4** Please write your analysis on Problem 4 here a)

Lemma 1: Let  $X_i = I_S[f_i]$ , then  $\mathbb{E}[X_i] = I[f]$  and by chebyshev's bound therefore:

$$Pr[|I_S[f_i] - I[f_i]| \geq \epsilon] \leq \frac{Var[X_i]}{\epsilon^2}$$

**Proof:** Let  $X_i = I_S[f_i] = \frac{1}{n} \sum_{j=1}^n V(f_i, z_j)$ , where  $z_j$  is the random samples/data and  $f_i$  is a non random (and fixed) function in  $\mathcal{H}$ . Then, if we take the expectation of  $X_i$  wrt to the distribution of  $z$  we get:

$$\mathbb{E}[X_i] = \mathbb{E}\left[\frac{1}{n} \sum_{j=1}^n V(f_i, z_j)\right] = \frac{1}{n} \sum_{j=1}^n \mathbb{E}[V(f_i, z_j)] = \mathbb{E}[V(f_i, z_j)] = I[f_i]$$

Thus, we can use chebyshev's and thus the following statement is true (concluding proof of lemma):

$$Pr[|I_S[f_i] - I[f_i]| \geq \epsilon] \leq \frac{Var[X_i]}{\epsilon^2}$$

□

Theorem: Given the conditions in the question, the the upper bound we are looking for is as follows:

$$Pr[\sup_{f \in \mathcal{H}} |I_S[f] - I[f]| \geq \epsilon] \leq \frac{N(c^2 - 2cM + M^2)^2}{n\epsilon^2}$$

**Proof:** If the largest difference between the empirical risk and the expected risk is larger than  $\epsilon$ , then that means that the defect (i.e. difference of empirical risk and generalization error) of one of the functions in this finite set is larger than  $\epsilon$ . i.e. At least one of the defects is larger than  $\epsilon$ . In equations it reads as follows:

$$Pr[\sup_{f \in \mathcal{H}} |I_S[f] - I[f]| \geq \epsilon] = Pr[\cup_{i=1}^N |I_S[f_i] - I[f_i]| \geq \epsilon]$$

by the union bound:

$$Pr[\sup_{f \in \mathcal{H}} |I_s[f] - I[f]| \geq \epsilon] \leq \sum_{i=1}^N Pr[|I_s[f_i] - I[f_i]| \geq \epsilon]$$

By Lemma 1 we know:

$$Pr[|I_s[f_i] - I[f_i]| \geq \epsilon] \leq \frac{Var[X_i]}{\epsilon^2}$$

Therefore using this upper bound we can further upper bound our equation above by:

$$\sum_{i=1}^N Pr[|I_s[f_i] - I[f_i]| \geq \epsilon] \leq \sum_{i=1}^N \frac{Var[X_i]}{\epsilon^2}$$

□

**Problem 5 (MATLAB)** Please write your analysis on Problem 5 here