1 Probability bound on approximation error E_{approx}

We assume that

- The labels b_i are binary.
- We use only one setting of hyper-parameters.
- the data-points in the validation set are identically and independently distributed (with replacement). This implies that $\mathbb{E}[E_{approx}] = 0$.
- The image of the loss function is in $[0, 1]$.

During class we claimed that

$$\mathbb{P}(E_{approx} > \epsilon) \leq 2\exp(-2\epsilon^2 t), \tag{1}$$

which can be proved by using Hoeffding's inequality (non-trivial).

We will now prove that if the above assumptions are true and

- we try ζ hyper-parameter settings,

then

$$\mathbb{P}(E_{approx} > \epsilon \text{ for any hyper-parameter}) \leq 2\zeta\exp(-2\epsilon^2 t).$$

Proof. Let's denote E_{approx} for a particular hyper-parameter setting $\tilde{\zeta}$ by $E_{approx}(\tilde{\zeta})$. If each run for each hyper-parameter setting is independent from other runs, then (1) applies for any hyper-parameter setting $\tilde{\zeta}$. However, (1) does not apply for $\hat{E}_{approx} := \min_{\tilde{\zeta}}$ by $E_{approx}(\tilde{\zeta})$. This is because the approximation error $\hat{E}_{approx}$ is biased, i.e., $\mathbb{E}[\hat{E}_{approx}] \neq 0$. This is because we chose the best hyper-parameter setting.

Let's consider the event $\hat{E}_{approx} > \epsilon$, then

$$\mathbb{P}(\hat{E}_{approx} > \epsilon) \leq \mathbb{P}(\underbrace{E_{approx}(\tilde{\zeta}) > \epsilon \text{ for any } \tilde{\zeta}}_{\text{union of } \zeta \text{ events}}).$$

We will use the *union bound*, i.e., for any events $\{A_1, A_2, \ldots, A_\zeta\}$ is that:

$$\mathbb{P}(A_1 \cup \cdots \cup A_\zeta) \leq \sum_{i=1}^{\zeta} \mathbb{P}(A_i).$$

(Prove this by noticing that $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq \mathbb{P}(A) + \mathbb{P}(B)$). Using the Union bound for our problem we get

$$\begin{aligned}
\mathbb{P}(\hat{E}_{approx} > \epsilon) &\leq \mathbb{P}(\underbrace{E_{approx}(\tilde{\zeta}) > \epsilon \text{ for any } \tilde{\zeta}}_{\text{union of } \zeta \text{ events}}) \\
&\leq \sum_{\tilde{\zeta}} \mathbb{P}(E_{approx}(\tilde{\zeta}) > \epsilon) \\
&\leq \sum_{\tilde{\zeta}} 2\exp(-2\epsilon^2 t) \text{ (using (1))} \\
&= 2\zeta\exp(-2\epsilon^2 t).
\end{aligned}$$

$\square$