

Solution 1:

- 1) The generalization error is the expected (future) performance of a learner \mathcal{I} trained on n_{train} observations with performance measure ρ . As any such performance estimate depends on the concrete sampling of $\mathcal{D}_{\text{test}}$ from \mathbb{P}_{xy} , we are interested in the limit of this expectation value, as $n_{\text{test}} \rightarrow \infty$.
- 2) One samples $\mathcal{D}_{\text{train}}$ of size $n_{\text{train}} = 100$ and $\mathcal{D}_{\text{test}}$ of size n_{test} K times from \mathbb{P}_{xy} (independently). Each time, the learner \mathcal{I} is trained on $\mathcal{D}_{\text{train},k}$, and the respective performance ρ is evaluated on $\mathcal{D}_{\text{test},k}$. For $K, n_{\text{test}} \rightarrow \infty$, the average performance $\frac{1}{K} \sum_{k=1}^K \rho(\mathbf{y}_{J_{\text{test},k}}, \mathbf{F}_{J_{\text{test},k}, \mathcal{I}(\mathcal{D}_{\text{train},k})})$ converges to $\text{GE}(\mathcal{I}, n_{\text{train}} = 100, \rho)$.
- 3) As n_{train} must be smaller than n , the estimator is a pessimistically biased estimator of $\text{GE}(\mathcal{I}, n, \rho)$, as we are not using all available data for training. In the context of regression tasks and performance measures MSE or MAE, pessimistic bias means: $\mathbb{E} \left[\widehat{\text{GE}}_{J_{\text{train}}, J_{\text{test}}}(\mathcal{I}, |J_{\text{train}}|, \rho) \right] > \text{GE}(\mathcal{I}, n, \rho)$
- 4) If one chooses a large n_{train} , n_{test} is small, and the estimator has a large variance.