Generalization Bounds via Validation

Thomas Walker

Notes on generalization bounds via validation methods. The theory behind these notes will be gathered from.

1 Notation

- $x \in \mathcal{X}$, input.
- $y \in \mathcal{Y}$, target.
- $\mathcal{L}$, loss function.
- $\mathcal{R}[f] = \mathbb{E}_{x,y\sim\mathbb{P}_{(\mathcal{X},\mathcal{Y})}}\left(\mathcal{L}(f(x), y)\right)$, expected risk of a function f. Where $\mathbb{P}_{(\mathcal{X},\mathcal{Y})}$ is the true distribution.
- $f_{\mathcal{A}(S)} : \mathcal{X} \to \mathcal{Y}$, function learnt by a learning algorithm $\mathcal{A}$ on training set $S := S_m := \{(x_1, y_1), \dots, (x_m, y_m)\}$
- $\mathcal{F}$, a characterized set of functions known as the hypothesis space.
- $\mathcal{L}_{\mathcal{F}} := \{g : g \in \mathcal{F}, g(x, y) = \mathcal{L}(f(x), y)\}$, the family of loss functions associated to $\mathcal{F}$.
- For vector v its dimension is d_v.

2 Setup

Machine learning aims to minimize $\mathcal{R}\left(f_{\mathcal{A}(S)}\right)$. However, this is **non-computable** as $\mathbb{P}_{(\mathcal{X},\mathcal{Y})}$ is unknown. Therefore, one minimizes the empirical risk

$$\mathcal{R}_S\left(f_{\mathcal{A}(S)}\right) = \frac{1}{|S|}\sum_{(x,y)\in S}\mathcal{L}(f_{\mathcal{A}(S)}(x), y),$$

where the generalization gap is given by $\mathcal{R}\left(f_{\mathcal{A}(S)}\right) - \mathcal{R}_S\left(f_{\mathcal{A}(S)}\right)$.

Affiliation Imperial College London
Correspondence thomas.walker21@imperial.ac.uk
Date July 2023

3 Generalization bounds via validation

The training-validation paradigm involves holding out a validation set to optimize model architecture. Giving rise to the hypothesis that *deep neural networks can obtain good generalization error by performing a model search on the validation set.*

Proposition 3.1 *Let* $S^{(\text{val})}_{m_{\text{val}}}$ *be a held-out validation set, where* $|S^{(\text{val})}_{m_{\text{val}}}| = m_{(\text{val})}$. *Assume that* $m_{(\text{val})}$ *is an* i.i.d *sample from* $\mathbb{P}_{(X,Y)}$. *Let* $\kappa_{f,i} = \mathcal{R}(f) - \mathcal{L}(f(x_i), y_i)$ *for* $(x_i, y_i) \in S^{(\text{val})}_{m_{\text{val}}}$. *Suppose that* $\mathbb{E}(\kappa^2_{f,i}) \leq \gamma^2$ *and* $|\kappa_{f,i}| \leq C$ *almost surely for all* $(f, i) \in \mathcal{F}_{\text{val}} \times \{1, \ldots, m_{\text{val}}$. *Then, for* $\delta \in (0, 1]$, *with probability* $1 - \delta$

$$\mathcal{R}(f) \leq \mathcal{R}_{S^{(\text{val})}_{m_{\text{val}}}}(f) + \frac{2C \log\left(\frac{|\mathcal{F}_{\text{val}}|}{\delta}\right)}{3m_{\text{val}}} + \sqrt{\frac{2\gamma^2 \log\left(\frac{|\mathcal{F}_{\text{val}}|}{\delta}\right)}{m_{\text{val}}}}$$

holds for all $f \in \mathcal{F}_{\text{val}}$.

Remarks 3.2

- $\mathcal{F}_{\text{val}}$ *is independent of* $S^{(\text{val})}_{m_{\text{val}}}$.
- *The bound is only dependent on the validation error on* $S^{(\text{val})}_{m_{\text{val}}}$.

Implementation 3.3