Iterative Approximate Cross Validation in High Dimensions

Sina Alsharif

School of Computer Science and Engineering, University of New South Wales

Thesis A (Term 2, 2023)

Section 1

Background

Background

To set out the notation used throughout this seminar, we can define the Empirical Risk Minimisation (ERM) framework to solve a supervised learning problem.

Background

To set out the notation used throughout this seminar, we can define the Empirical Risk Minimisation (ERM) framework to solve a supervised learning problem.

General Problem Setting

- Input space $\mathcal{X}$ and an output space $\mathcal{Y}$.
- Generated by a *true* distribution $P(\mathcal{X}, \mathcal{Y})$.
- Aim is to find a mapping $h : \mathcal{X} \to \mathcal{Y}$ (called a hypothesis).
- Denote all possible combinations of input and output space as $\mathcal{D} = \{(X, Y) \in (\mathcal{X}, \mathcal{Y})\}$ (called a hypothesis).

Risk

To measure the error (or loss) we make on a data point, define a function $\ell(h; D_i)$,

$$\ell(h; D) = \sum_{i=1}^{n} \ell(h; D_i)$$

as the loss for the dataset. Examples of ℓ are 0-1 loss (for classification) and a squared error (for regression).

Risk

To measure the error (or loss) we make on a data point, define a function $\ell(h; D_i)$,

$$\ell(h; D) = \sum_{i=1}^{n} \ell(h; D_i)$$

as the loss for the dataset. Examples of ℓ are 0-1 loss (for classification) and a squared error (for regression).

Risk

Define the **risk** of a hypothesis for the complete space of inputs and outputs as,

$$R(h) = \mathbb{E}_{\mathcal{X},\mathcal{Y}}[\ell(h; \mathcal{D})]$$

The optimal hypothesis for the data is,

$$h_{\mathcal{H}} = \arg\min_{h \in \mathcal{H}} R(h)$$

Empirical Risk

As we cannot measure **true** risk, we seek an approximation using the observed data D.

Empirical Risk

We define **empirical risk** of a hypothesis given data D as,

$$R_{\text{emp}}(h; D) = \frac{1}{n} \sum_{i=1}^{n} \ell(h; D_i)$$

The optimal hypothesis *given observed data* D is,

$$h_D = \arg\min_{h \in \mathcal{H}} R_{\text{emp}}(h; D)$$

where $\mathcal{H}$ is a hypothesis space which a *learning algorithm* picks a hypothesis from.

Section 2

Literature Review

Section 3

Preliminary Work

Section 4

Future Plans