Unbiased estimates for linear regression via volume sampling

Michał Dereziński  Manfred K. Warmuth

Linear regression

\[ L(w) = \sum_{i=1}^{n} (x_i^T w - y_i)^2 \]

\[ w^* = \text{argmin}_w L(w) \]

Subsampling for linear regression

**Given:** \( n \) points \( x_i \in \mathbb{R}^d \) with hidden labels \( y_i \in \mathbb{R} \)

**Goal:** Minimize loss \( L(w) = \sum (x_i^T w - y_i)^2 \) over all \( n \) points

Select \( S = \{4, 6, 9\} \)

Receive \( y_1, y_6, y_9 \)

Simple strategy: Solve the subproblem, \( w^*(S) = X^S y^S \)

Volume sampling

\[ S \subseteq \{1...n\} \text{ chosen w.p.} \]

\[ \sim \text{ squared volume of parallelepiped} \]

spanned by the \( \{x_i : i \in S\} \)

Distribution over all \( d \)-element subsets \( S \):

\[ P(S) = \frac{\det(X_S X_S^T)}{Z} \]

Normalization factor obtained via Cauchy-Binet formula:

\[ Z = \sum_{|S|=d} \det(X_S X_S^T) = \det(X X^T) \]

Unbiased estimator for pseudo-inverse \( X^+ \)

**Key trick:** To each subset \( S \) assign a formula \( F(S) \) at

\[ F(S) = \sum_{i \in S} P(S \mid S_i) F(S_{-i}) \]

Then:

\[ \mathbb{E}_S [F(S)] = F(\{1...n\}) \]

**Expectation formulas for \( (X_S)^+ \):**

1. \( \mathbb{E}[X_S]^+] = X^+ \)  (unbiasedness)
2. \( \mathbb{E}[X_S X_S^T]^{-1} = \frac{n-d}{n} \frac{(X X^T)^{-1}}{(X^T X)^{-1}} \) (variance bound)

**Corollary:** \( \mathbb{E}[w^*(S)] = \mathbb{E}[(X_S)^+ y] = X^+ y = w^* \)

Averaging unbiased estimators

Let \( \bar{y}(S) = X^T w^*(S) \). If \( w^*(S) \) is unbiased \( \mathbb{E}[w^*(S)] = w^* \), then:

\[ \mathbb{E}[L(w^*(S))] \leq (1+c) L(w^*) \]

Take average of \( k \) i.i.d. samples of size \( s \): \( \bar{w} = \frac{1}{k} \sum_{s \in S} \bar{y}(S) \)

\[ \mathbb{E}[L(\bar{w}(S))] \leq \left(1 + \frac{c}{\epsilon}\right) L(w^*) \]

With size \( d \) volume sampling, we need \( d^2/\epsilon \) labels. Is \( d/\epsilon \) possible?

**Open:** Is there unbiased estimator with \( s = O(d) \) and \( c = O(1) \)?