
Journal

Simon Hugo
hugo.simon@telecom-paris.fr

1 Introduction

Let $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ be a set of n datapoints. Let Θ be a space of parameters, and θ an element of Θ . We consider cost functions of the form:

$$L(\theta) = \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$$

Let $\mathcal{S} = \{\mathbf{x}_{s_1}, \dots, \mathbf{x}_{s_m}\}$ be a subset of \mathcal{X} (possibly with repetitions). To each element $\mathbf{x}_s \in \mathcal{S}$, associate a weight $\omega(\mathbf{x}_s) \in \mathbb{R}^+$. Define the estimated cost associated to the weighted subset \mathcal{S} as:

$$\hat{L}(\theta) = \sum_{\mathbf{x}_s \in \mathcal{S}} \omega(\mathbf{x}_s) f(\mathbf{x}_s, \theta).$$

Definition 1.1 (Coreset). Let $\varepsilon \in]0, 1[$. The weighted subset \mathcal{S} is a ε -coreset for L if, for any parameter θ , the estimated cost is equal to the exact cost up to a relative error:

$$\forall \theta \in \Theta \quad \left| \frac{\hat{L}(\theta)}{L(\theta)} - 1 \right| \leq \varepsilon$$

An important consequence of the coreset property is the following

$$(1 - \varepsilon)L(\theta^{\text{opt}}) \leq (1 - \varepsilon)L(\hat{\theta}^{\text{opt}}) \leq \hat{L}(\hat{\theta}^{\text{opt}}) \leq \hat{L}(\theta^{\text{opt}}) \leq (1 + \varepsilon)L(\theta^{\text{opt}})$$

Definition 1.2 (Sensitivity). The sensitivity σ_i of a datapoint \mathbf{x}_i and the total sensitivity \mathfrak{S} of \mathcal{X} are

$$\begin{cases} \sigma_i = \sup_{\theta \in \Theta} \frac{f(\mathbf{x}_i, \theta)}{L(\mathcal{X}, \theta)} & \in [0, 1] \\ \mathfrak{S} = \sum_{i=1}^n \sigma_i \end{cases}$$

2 Variance argument

2.1 Multinomial case

Multinomial case $\mathcal{S} \sim \mathcal{M}(m, q)$ i.e. m independent categorical sampling where $\mathbb{P}(x_i) = q(x_i)$

$$\text{Var}[\hat{L}(\theta)] = \frac{1}{m} \text{Var} \left[\frac{f_{\theta}(x)}{q(x)} \right] = \frac{1}{m} \sum_{x \in \mathcal{X}} \frac{f_{\theta}(x)^2}{q(x)} - \frac{1}{m} L(\theta)^2$$

For any query $\theta \in \Theta$, the variance is minimized by

$$q_{\theta}(x) = \frac{f_{\theta}(x)}{\sum_{x' \in \mathcal{X}} f_{\theta}(x')},$$

2.2 DPP case

DPP case where $\mathcal{S} \sim \mathcal{DPP}(K)$, $\pi_i := K_{ii}$. We have

$$\text{Var}[\hat{L}(\theta)] = \sum_{i,j} \mathbb{E}[\varepsilon_i \varepsilon_j] \frac{f_\theta(x_i) f_\theta(x_j)}{\pi_i \pi_j} - L(\theta)^2 \quad \text{with} \quad \mathbb{E}[\varepsilon_i \varepsilon_j] = \begin{cases} \det(K_{ij}) = \pi_i \pi_j - K_{ij}^2, & \text{if } i \neq j \\ \mathbb{E}[\varepsilon_i] = \pi_i, & \text{if } i = j \end{cases}$$

Introducing $\Pi = \text{diag}(\pi)$ and $\tilde{K} = \Pi^{-1} K \Pi^{-1}$, we can rewrite

$$\mathbb{V}\text{ar}[\hat{L}(\theta)] = \sum_i \left(\frac{1}{\pi_i} - 1 \right) f_\theta(x_i)^2 - \sum_{i \neq j} \frac{K_{ij}^2}{\pi_i \pi_j} f_\theta(x_i) f_\theta(x_j) = f_\theta^\top (\Pi^{-1} - \tilde{K}) f_\theta$$

For a Bernoulli process where $\mathbb{P}(x_i \in \mathcal{S}) = \pi_i$ independently, $K = \Pi$ then $\tilde{K} = I$. The DPP variance beats uniformly the Bernoulli process variance if \tilde{K} dominates the identity i.e.

$$\forall f_\theta, \mathbb{V}\text{ar}[\hat{L}_K(\theta)] < \mathbb{V}\text{ar}[\hat{L}_\Pi(\theta)] \iff \tilde{K} > I$$

But \tilde{K} is a symmetric positive definite matrix and by Hadamard inequality $\det(\tilde{K}) \leq \prod_i \tilde{K}_{ii} = 1$. Therefore at least one of its eigenvalue is lower than 1, hence $\tilde{K} \not> I$.