数据科学导论第十五次课

Motivation

- airline data.
    - $n = 120$ million, $p = 29$.
    - Each observation corresponds to a piece of flight information within the United States.
    - The data occupies about 12G of space on the hard disk.
- Census income data set.
    - $n = 48,842$.
    - want to conduct a classification analysis using residents' information such as age, work class, education and etc to predict whether the residents are high income residents, i.e., those with annal income more than $50K, or not.

Statistical analysis challenge for Big Data

- Extraordinary size of data.
- Limited computation resource.
- Data reduction: The focus is on reducing the size of n.
- Selecting a subdata suitable for limited computing resource.
- Tradeoffs Between Computational Costs and Statistical Efficiency.

1. Subsampling for linear model
2. Subsampling for Logistic model

Big Data Linear Regression

- Linear model:

$$\mathbf{y} = \boldsymbol{\beta}_0 \mathbf{1} + \boldsymbol{Z}\boldsymbol{\beta}_1 + \boldsymbol{\epsilon} = \boldsymbol{X}^\top \boldsymbol{\beta} + \boldsymbol{\epsilon},$$

  where $\boldsymbol{Z}$ is an $n \times p$ matrix, $\boldsymbol{X} = [\mathbf{1}, \boldsymbol{Z}]$, y_i's are uncorrelated given $\boldsymbol{Z}$, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^{\mathbf{2}}\boldsymbol{I})$.

- OLS (Ordinary Least Squares):

$$\hat{\boldsymbol{\beta}}_{OLS} = \mathop{\arg\min}_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\mathbf{y} - \boldsymbol{X}\boldsymbol{\beta}\|^{\mathbf{2}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-\mathbf{1}} \boldsymbol{X}^\top \boldsymbol{y}.$$

Computational Cost

- Inference under linear regression has a computational complexity of $O(np^2)$, which can be problematic for large n.
- Wisely chosen subdata can be used so that useful conclusions can be obtained with limited computational resources.
- If subdata size is k, the overall computational complexity is $O(kp^2 + ?)$, where ? depends on the computational complexity of the algorithm for selecting the subdata.

subsampling-based methods

- Assign each data point $(\boldsymbol{x}_i, y_i)$ a proper sampling probability π_i, $\sum_{i=1}^n \pi_i = 1$, and sample $k(k \ll n)$ data points with replacement according to $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$, and do statistical inference using the subdata.
- How to define π_i so that the data points that really affect the result have a greater probability of being collected?
- How to analyze the k data points to get a estimator close enough to the estimator based on the full data?

Sampling strategies

- Uniform sampling (UNIF): $\pi_i = 1/n$.
- Leverage-based sampling (LEV): $\pi_i = h_{ii}/\sum_{i=1}^{n} h_{ii}$ where h_{ii} is the leverage score of the i^{th} data point.
- Information-Based Optimal Subdata Selecion (IBOSS).

Leverage-based Subsampling

- $\hat{\mathbf{y}} = \boldsymbol{H}\boldsymbol{y}, H = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{\mathbf{-1}}\boldsymbol{X}^\top.$
- $h_{ii} = \boldsymbol{x}_i^\top(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{x}_i$. Generally, the greater the leverage score, the more important the data point is.
- $0 \leq h_{ii} \leq 1$: First, note that $\boldsymbol{H}$ is an idempotent matrix satisfying $\boldsymbol{H} = \boldsymbol{H}^{\mathbf{2}}$, then $h_{ii} = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2 \Rightarrow h_{ii} > h_{ii}^2$.
- Define the i^{th} regression residual $e_i = y_i - \hat{y}_i$, we have $\mathrm{Var}(e_i) = (1 - h_{ii})\sigma^2$. In other words, the greater the score, the closer $\hat{y}_i$ will be to the y_i.

How to analyze the subdata?

- Use these k data points directly for linear regression.
- Reweight these k data points.

Formulation

- sampling matrix $\boldsymbol{S}_{k\times n}^{\top}$:
    - $\boldsymbol{S}^{\top} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k)^{\top}$ where $\boldsymbol{e}_j \in \mathbb{R}^n$ be a unit vector with the j^{th} element being 1 and 0 otherwise.
    - if the r^{th} data unit (or observation) in the original data set is chosen in the i^{th} random trial, then the i^{th} row of $\boldsymbol{S}^{\top}$ equals $\boldsymbol{e}_r$;
    - The subsample: $(\boldsymbol{X}^*, \boldsymbol{y}^*) = (\boldsymbol{S}^{\top}\boldsymbol{X}, \boldsymbol{S}^{\top}\boldsymbol{y})$.
- rescaling (reweighting) matrix $\boldsymbol{D}$:
    - an $k \times k$ diagonal matrix whose i^{th} diagonal element equals $1/\sqrt{k\pi_r}$ if the r^{th} data point is chosen in the i^{th} random trial.
    - every diagonal element of $\boldsymbol{D}$ equals $\sqrt{n/k}$ for uniform sampling.

SubsampleLS

- $$\tilde{\beta}_W = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min} \|\boldsymbol{D}\boldsymbol{S}^\top \mathbf{y} - \boldsymbol{D}\boldsymbol{S}^\top \boldsymbol{X}\beta\|^{\mathbf{2}} = (\boldsymbol{X}^\top \boldsymbol{S}\boldsymbol{D}^{\mathbf{2}}\boldsymbol{S}^\top \boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^\top \boldsymbol{S}\boldsymbol{D}^{\mathbf{2}}\boldsymbol{S}^\top \boldsymbol{y}.$$

- $$\boldsymbol{W} = \boldsymbol{S}\boldsymbol{D}^2\boldsymbol{S}^\top, \;\; \hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{y}.$$

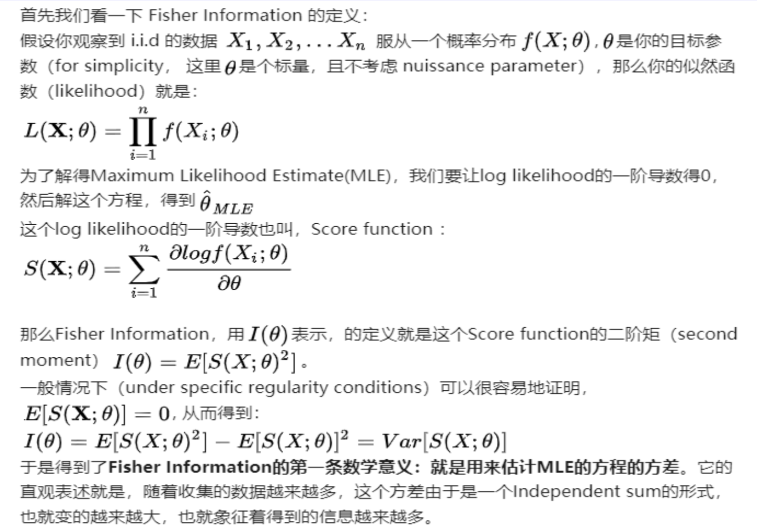
Random sampling approach

- Uniform Sampling Estimator (UNIF): uniform subsampling and weighted LS estimation.
- Basic Leveraging Estimator (LEV): exact leverage-based sampling and weighted LS estimation,
- Shrinkage Leveraging Estimator (SLEV): shrinkage leverage-based sampling and weighted LS estimation: $\pi_i = \alpha\pi_i^{Lev} + (1-\alpha)\pi_i^{Unif}, \alpha \in (0,1)$.
- Unweighted Leveraging Estimator (LEVUNW): leverage-based sampling and unweighted LS estimation.

Information-based subsampling

- How to preserve the majority information contained in the full data ?

Fisher information



首先我们看一下 Fisher Information 的定义：

假设你观察到 i.i.d 的数据 $X_1, X_2, ... X_n$ 服从一个概率分布 $f(X;\theta)$，θ是你的目标参数（for simplicity，这里θ是个标量，且不考虑 nuissance parameter），那么你的似然函数（likelihood）就是：

$$L(\mathbf{X};\theta)=\prod_{i=1}^{n} f(X_i;\theta)$$

为了解得Maximum Likelihood Estimate(MLE)，我们要让log likelihood的一阶导数得0，然后解这个方程，得到 $\hat{\theta}_{MLE}$

这个log likelihood的一阶导数也叫，Score function：

$$S(\mathbf{X};\theta)=\sum_{i=1}^{n} \frac{\partial log f(X_i;\theta)}{\partial\theta}$$

那么Fisher Information，用$I(\theta)$表示，的定义就是这个Score function的二阶矩（second moment）$I(\theta)=E[S(X;\theta)^2]$。

一般情况下（under specific regularity conditions）可以很容易地证明，

$E[S(\mathbf{X};\theta)]=0$，从而得到：

$I(\theta)=E[S(X;\theta)^2]-E[S(X;\theta)]^2=Var[S(X;\theta)]$

于是得到了**Fisher Information的第一条数学意义：就是用来估计MLE的方程的方差。**它的直观表述就是，随着收集的数据越来越多，这个方差由于是一个Independent sum的形式，也就变的越来越大，也就象征着得到的信息越来越多。

IBOSS

- Rather than sampling, we select subdata so as to maximize the Fisher information matrix.
- For linear regression, under normality and taking $\sigma^2 = 1$ for simplicity, the information matrix for $\boldsymbol{\beta}$ with subdata is

$$\boldsymbol{M}(\boldsymbol{\delta}) = \sum_{i=1}^{n} \delta_i x_i x_i^\top = \boldsymbol{X}^\top \boldsymbol{\Delta} \boldsymbol{X}.$$

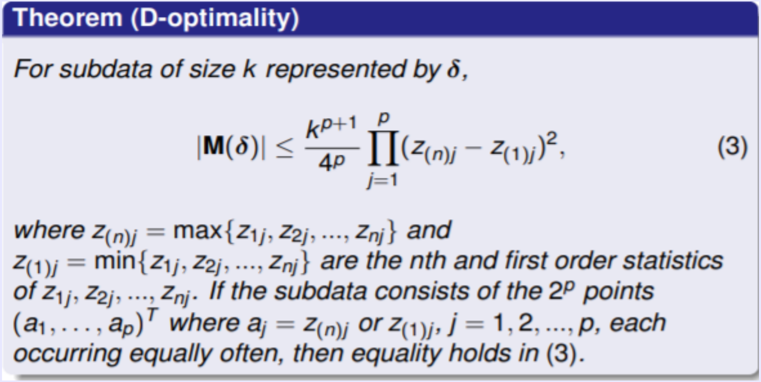
with δ_i an "inclusion" indicator, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$ and $\boldsymbol{\Delta} = \mathrm{diag}\{\boldsymbol{\delta}\}$.

Optimal Design of Experiments

- As in optimal design of experiments (DOE), we could maximize a function of the information matrix.
- D-optimality: Maximize the determinant of the information matrix.
- Find $\boldsymbol{\delta}$, subject to $\sum_{i=1}^{n} \delta_i = k$, that maximizes $\det(M(\boldsymbol{\delta}))$.
- Need a computationally efficient algorithm to find, approximately, an optimal $\boldsymbol{\delta}$.

D-optimal design under approximate design theory

An upper bound for $\det M(\delta)$

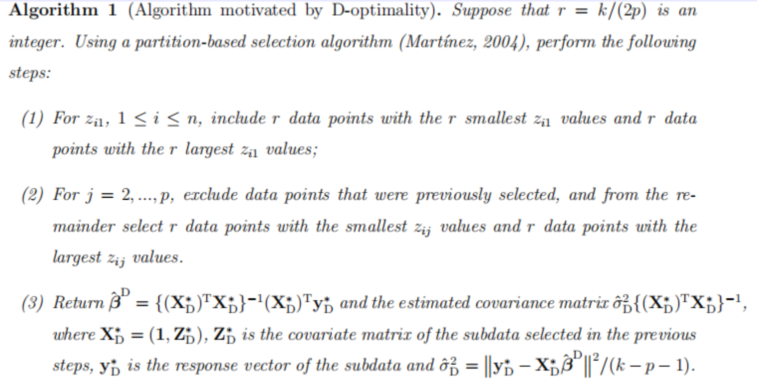


Theorem (D-optimality)

For subdata of size k represented by δ,

$$|\mathbf{M}(\delta)| \leq \frac{k^{p+1}}{4^p} \prod_{j=1}^{p} (z_{(n)j} - z_{(1)j})^2, \tag{3}$$

where $z_{(n)j} = \max\{z_{1j}, z_{2j}, ..., z_{nj}\}$ and $z_{(1)j} = \min\{z_{1j}, z_{2j}, ..., z_{nj}\}$ are the nth and first order statistics of $z_{1j}, z_{2j}, ..., z_{nj}$. If the subdata consists of the 2^p points $(a_1, \ldots, a_p)^T$ where $a_j = z_{(n)j}$ or $z_{(1)j}$, $j = 1, 2, ..., p$, each occurring equally often, then equality holds in (3).



Algorithm 1 (Algorithm motivated by D-optimality). *Suppose that $r = k/(2p)$ is an integer. Using a partition-based selection algorithm (Martínez, 2004), perform the following steps:*

(1) For z_{i1}, $1 \leq i \leq n$, include r data points with the r smallest z_{i1} values and r data points with the r largest z_{i1} values;

(2) For $j = 2, ..., p$, exclude data points that were previously selected, and from the remainder select r data points with the smallest z_{ij} values and r data points with the largest z_{ij} values.

(3) Return $\hat{\boldsymbol{\beta}}^{\mathrm{D}} = \{(\mathbf{X}_{\mathrm{D}}^{*})^{\mathrm{T}}\mathbf{X}_{\mathrm{D}}^{*}\}^{-1}(\mathbf{X}_{\mathrm{D}}^{*})^{\mathrm{T}}\mathbf{y}_{\mathrm{D}}^{*}$ and the estimated covariance matrix $\hat{\sigma}_{\mathrm{D}}^{2}\{(\mathbf{X}_{\mathrm{D}}^{*})^{\mathrm{T}}\mathbf{X}_{\mathrm{D}}^{*}\}^{-1}$, where $\mathbf{X}_{\mathrm{D}}^{*} = (\mathbf{1}, \mathbf{Z}_{\mathrm{D}}^{*})$, $\mathbf{Z}_{\mathrm{D}}^{*}$ is the covariate matrix of the subdata selected in the previous steps, $\mathbf{y}_{\mathrm{D}}^{*}$ is the response vector of the subdata and $\hat{\sigma}_{\mathrm{D}}^{2} = \left\|\mathbf{y}_{\mathrm{D}}^{*} - \mathbf{X}_{\mathrm{D}}^{*}\hat{\boldsymbol{\beta}}^{\mathrm{D}}\right\|^{2}/(k-p-1)$.

Algorithm for D-optimality

- To maximize $\det(M(\delta))$, we need to include points with large and small covariate values.
- For a fixed subdata size k, using a partition-based selection algorithm, for $j = 1, \ldots, p$, select the $k/(2p)$ largest and smallest z_{ij}-values, and include these points in the subdata
- $\hat{\boldsymbol{\beta}}^D = (\boldsymbol{X}^\top \boldsymbol{\Delta} \boldsymbol{X})^{-\mathbf{1}} \boldsymbol{X}^\top \boldsymbol{\Delta} \boldsymbol{y}$.
- Computational complexity for selection $O(np)$; Overall computational complexity $O(kp^2 + np)$, or $O(np)$ if $n > kp$.
- Can do this one covariate at a time or in parallel.

Theoretical property

- IBOSS can be used no matter what the distribution of the covariates is.
- Let $z_1, \ldots, z_n$ be iid, and consider 3 scenarios:
  - $z_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
  - $z_i \sim \text{Lognormal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
  - $z_i \sim t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- For all scenarios, $\text{Var}(\hat{\beta}_0^D|Z)$ is proportional to $1/k$ when $n \to \infty$.
- Elements of $\text{Var}(\hat{\beta}_1^D|Z)$ converge to 0 when $n \to \infty$ in all cases. (even though the subdata size k is fixed)
- Similar results typically do not hold for random subsampling methods.

Extension to large p

- Penalized likelihood estimators: LASSO etc.
- For linear model,

$$\hat{\boldsymbol{\beta}}_{LASSO} = \arg\min_{\boldsymbol{\beta}} \left\{ \|y - \boldsymbol{X}\boldsymbol{\beta}\|^2/n + \lambda\|\boldsymbol{\beta}\|_1 \right\}.$$

- want to develop subset selection method
- D-optimality Motivated algorithm??

1. Subsampling for linear model
2. Subsampling for Logistic model

Logistic regression

- Dependent variable is 0-1 type data.
- Sigmoid funcgtion:

$$s(x) = \frac{\exp(x)}{1 + \exp(x)}.$$

- $$\mathrm{P}(Y_i = 1|X_i) = p_i(\beta) = \frac{\exp(X_i^\top \beta)}{1 + \exp(X_i^\top \beta)}.$$

MLE

- The joint likelihood is:

$$L(\beta) = \prod_{i=1}^{n} \mathrm{P}(Y_i \mid X_i) = \prod_{i=1}^{n} p_i(\beta)^{Y_i}(1 - p_i(\beta))^{1-Y_i}$$

-

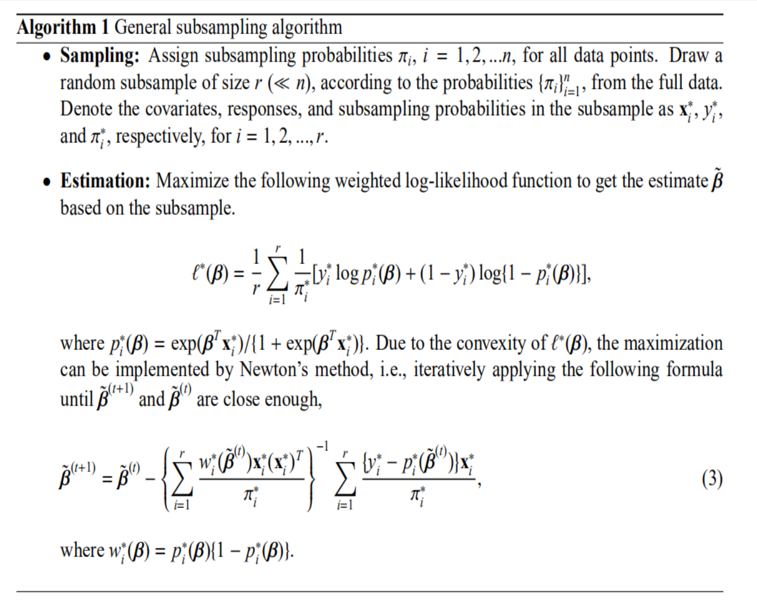
$$\hat{\beta}_{MLE} = \arg\max_{\beta} l(\beta) = \arg\max_{\beta} \sum_{i=1}^{n} [Y_i \log(p_i(\beta)) + (1 - Y_i)\log(1 - p_i(\beta))].$$

- Newton iteration method:

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} - \left\{ \sum_{i=1}^{n} w_i(\hat{\beta}^{(t)}) x_i x_i^{\top} \right\}^{-1} \frac{\partial l(\hat{\beta}^{(t)})}{\partial \beta} \text{where.}$$

- Computation cost is $O(\zeta n p^2)$, ζ is the iteration number.

Levraging type algorithm



Algorithm 1 General subsampling algorithm

- **Sampling:** Assign subsampling probabilities π_i, $i = 1, 2, ...n$, for all data points. Draw a random subsample of size r ($\ll n$), according to the probabilities $\{\pi_i\}_{i=1}^{n}$, from the full data. Denote the covariates, responses, and subsampling probabilities in the subsample as $\mathbf{x}_i^*$, y_i^*, and π_i^*, respectively, for $i = 1, 2, ..., r$.

- **Estimation:** Maximize the following weighted log-likelihood function to get the estimate $\tilde{\boldsymbol{\beta}}$ based on the subsample.

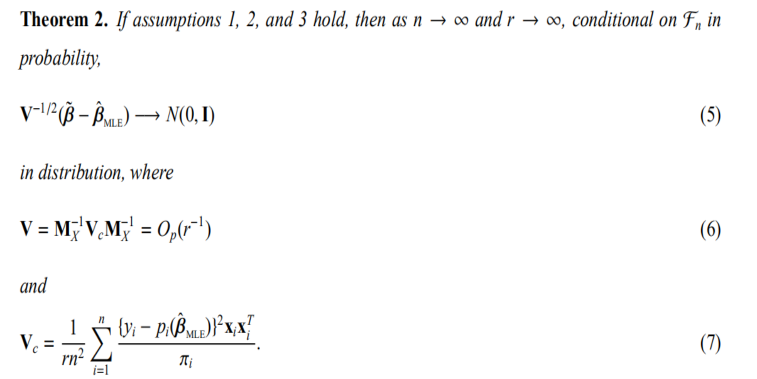
$$\ell^*(\boldsymbol{\beta}) = \frac{1}{r}\sum_{i=1}^{r}\frac{1}{\pi_i^*}[y_i^* \log p_i^*(\boldsymbol{\beta}) + (1 - y_i^*)\log\{1 - p_i^*(\boldsymbol{\beta})\}],$$

where $p_i^*(\boldsymbol{\beta}) = \exp(\boldsymbol{\beta}^T\mathbf{x}_i^*)/\{1 + \exp(\boldsymbol{\beta}^T\mathbf{x}_i^*)\}$. Due to the convexity of $\ell^*(\boldsymbol{\beta})$, the maximization can be implemented by Newton's method, i.e., iteratively applying the following formula until $\tilde{\boldsymbol{\beta}}^{(t+1)}$ and $\tilde{\boldsymbol{\beta}}^{(t)}$ are close enough,

$$\tilde{\boldsymbol{\beta}}^{(t+1)} = \tilde{\boldsymbol{\beta}}^{(t)} - \left\{\sum_{i=1}^{r}\frac{w_i^*(\tilde{\boldsymbol{\beta}}^{(t)})\mathbf{x}_i^*(\mathbf{x}_i^*)^T}{\pi_i^*}\right\}^{-1}\sum_{i=1}^{r}\frac{\{y_i^* - p_i^*(\tilde{\boldsymbol{\beta}}^{(t)})\}\mathbf{x}_i^*}{\pi_i^*}, \tag{3}$$

where $w_i^*(\boldsymbol{\beta}) = p_i^*(\boldsymbol{\beta})\{1 - p_i^*(\boldsymbol{\beta})\}$.

Asymptotically normal



Theorem 2. *If assumptions 1, 2, and 3 hold, then as $n \to \infty$ and $r \to \infty$, conditional on $\mathcal{F}_n$ in probability,*

$$\mathbf{V}^{-1/2}(\tilde{\beta} - \hat{\beta}_{\mathrm{MLE}}) \longrightarrow N(0, \mathbf{I}) \tag{5}$$

in distribution, where

$$\mathbf{V} = \mathbf{M}_X^{-1}\mathbf{V}_c\mathbf{M}_X^{-1} = O_p(r^{-1}) \tag{6}$$

and

$$\mathbf{V}_c = \frac{1}{rn^2}\sum_{i=1}^{n}\frac{\{y_i - p_i(\hat{\beta}_{\mathrm{MLE}})\}^2\mathbf{x}_i\mathbf{x}_i^T}{\pi_i}. \tag{7}$$

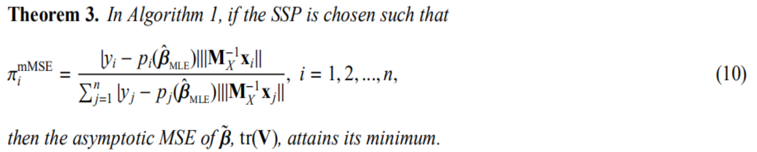
Asymptotic properties

- The subdata size is r.
- $\tilde{\boldsymbol{\beta}}$ is consistent to $\hat{\boldsymbol{\beta}}_{MLE}$ given the full sample and the convergence rate is $O(r^{1/2})$.
- $(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{MLE}) \mid \mathcal{F}_n \overset{a}{\rightsquigarrow} \boldsymbol{u}$ where $\boldsymbol{u}$ is a normal random variable with distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{V})$.
- Under some condition,

$$\mathbb{E}\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{MLE}\|^2 \mid \mathcal{F}_n\right) \text{ is close to } \mathbb{E}(\|\boldsymbol{u}\|^2 \mid \mathcal{F}_n).$$

- Minimize the asymptotic MSE to obtain the sampling probability.

OSMAC



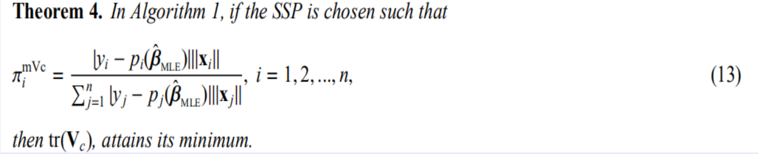
Theorem 3. *In Algorithm 1, if the SSP is chosen such that*

$$\pi_i^{\text{mMSE}} = \frac{|y_i - p_i(\hat{\boldsymbol{\beta}}_{\text{MLE}})| \|\mathbf{M}_X^{-1}\mathbf{x}_i\|}{\sum_{j=1}^{n} |y_j - p_j(\hat{\boldsymbol{\beta}}_{\text{MLE}})| \|\mathbf{M}_X^{-1}\mathbf{x}_j\|}, \quad i = 1, 2, \ldots, n, \tag{10}$$

then the asymptotic MSE of $\tilde{\boldsymbol{\beta}}$, $\text{tr}(\mathbf{V})$, *attains its minimum.*

Since this optimal subsampling procedure is motivated from the A-optimality criterion, it is called OSMAC.

- For two positive definite matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, $\boldsymbol{A}_1 \leq \boldsymbol{A}_2$ if and only if $\boldsymbol{A}_1 - \boldsymbol{A}_2$ is a non-negative definite matrix.
- $\boldsymbol{V} = \boldsymbol{M_x^{-1} V_c M_x^{-1}}$ depends on π through $\boldsymbol{V}_c$ and $\boldsymbol{M}_x$ does not depend on π, for given $\pi_{(1)}$ and $\pi_{(2)}$, $\boldsymbol{V}(\pi_{(1)}) \leq \boldsymbol{V}(\pi_{(2)})$ if and only if $\boldsymbol{V}_c(\pi_{(1)}) \leq \boldsymbol{V}_c(\pi_{(2)})$.
- minimize $\mathrm{tr}(\boldsymbol{V_c})$ instead of minimizing $\mathrm{tr}(\boldsymbol{V})$.



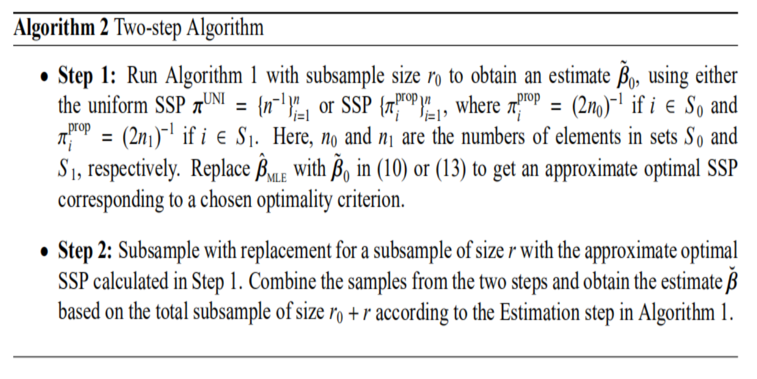
Theorem 4. *In Algorithm 1, if the SSP is chosen such that*

$$\pi_i^{\mathrm{mVc}} = \frac{|y_i - p_i(\hat{\boldsymbol{\beta}}_{\mathrm{MLE}})|\|\mathbf{x}_i\|}{\sum_{j=1}^{n} |y_j - p_j(\hat{\boldsymbol{\beta}}_{\mathrm{MLE}})|\|\mathbf{x}_j\|}, \quad i = 1, 2, \ldots, n, \tag{13}$$

then $\mathrm{tr}(\mathbf{V}_c)$, *attains its minimum.*

- Let $S_0 = \{i : y_i = 0\}$ and $S_1 = \{i : y_i = 1\}$, the effect of $p_i(\hat{\beta}_{MLE})$ on π_i^{mMSE} is positive for the S_0 set while the effect is negative for the S_1 set.
- The optimal subsampling approach is more likely to select data points with smaller $p_i(\hat{\beta}_{MLE})$'s when y_i's are 1 and data points with larger $p_i(\hat{\beta}_{MLE})$'s when y_i's are 0.
- Intuitively, it attempts to give preferences to data points that are more likely to be mis-classified.

Two-stage framework



Algorithm 2 Two-step Algorithm

- **Step 1:** Run Algorithm 1 with subsample size r_0 to obtain an estimate $\tilde{\boldsymbol{\beta}}_0$, using either the uniform SSP $\boldsymbol{\pi}^{\mathrm{UNI}} = \{n^{-1}\}_{i=1}^{n}$ or SSP $\{\pi_i^{\mathrm{prop}}\}_{i=1}^{n}$, where $\pi_i^{\mathrm{prop}} = (2n_0)^{-1}$ if $i \in S_0$ and $\pi_i^{\mathrm{prop}} = (2n_1)^{-1}$ if $i \in S_1$. Here, n_0 and n_1 are the numbers of elements in sets S_0 and S_1, respectively. Replace $\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ with $\tilde{\boldsymbol{\beta}}_0$ in (10) or (13) to get an approximate optimal SSP corresponding to a chosen optimality criterion.

- **Step 2:** Subsample with replacement for a subsample of size r with the approximate optimal SSP calculated in Step 1. Combine the samples from the two steps and obtain the estimate $\breve{\boldsymbol{\beta}}$ based on the total subsample of size $r_0 + r$ according to the Estimation step in Algorithm 1.

New Subsampling Algorithm

- Can we subset data via D-optimality similar to linear model?
- Yes!

References

[1] Ma, P., Mahoney, M. W., and Yu, B. (2015), A statistical perspective on algorithmic leveraging, *The Journal of Machine Learning Research*, 16, 861–911.

[2] Wang, H., Yang, M., and Stufken, J. (2019), Information-based optimal subdata selection for big data linear regression, *Journal of the American Statistical Association*, 114, 393–405.

[3] Wang, H., Zhu, R., and Ma, P. (2018), Optimal subsampling for large sample logistic regression, *Journal of the American Statistical Association*, 113, 829–844.

The end!

Thank you!