SLIDS Homework 2

Dave Zachariah

February 2, 2023

- Solution proposals are individual.
- Each solution must be reproducible by your peer. Code should be added to the appendix.
- The solution to each subproblem yield 0 to 2 points.

1

Let us revisit the ship localization problem from HW1. Suppose we are located at coordinates

$$\mathbf{a} = \begin{bmatrix} 50 \\ 100 \end{bmatrix}$$

We want to infer our distance to the ship

$$\tau_\circ = \|\boldsymbol{\theta}_\circ - \mathbf{a}\|$$

a) We first construct a confidence set $\mathcal{T}_\alpha^n$ based on the ERM by forming the estimate of the asymptotic variance v_n (see lecture slides).

   Show that

$$v_n = \frac{1}{\|\widehat{\boldsymbol{\theta}}_n - \mathbf{a}\|^2} (\widehat{\boldsymbol{\theta}}_n - \mathbf{a})^\top \, \mathbb{E}_n \left[(\widehat{\boldsymbol{\theta}}_n - \mathbf{z})(\widehat{\boldsymbol{\theta}}_n - \mathbf{z})^\top \right] (\widehat{\boldsymbol{\theta}}_n - \mathbf{a})$$

b) We construct intervals

$$\mathcal{T}_\alpha^n = [\tau_{\min}, \tau_{\max}]$$

   using the ERM-based method, with an intended coverage of $1-\alpha$ corresponding to 90%, 95% and 99%, respectively (aka. confidence levels):. Report on the distance uncertainties when obtaining $n = 100$ and $n = 1000$ samples.

   For the data generating process, use the Gaussian training distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ from HW1.

c) Since you are generating the data, the value of $\tau_\circ = \|\boldsymbol{\theta}_\circ - \mathbf{a}\|$ is known here. Let us now evaluate the accuracy of the constructed confidence intervals.

   Repeat the experiment above, by drawing M different sets of training data $\mathbf{z}^n$. For each set, construct a 95%-confidence interval:

$$\mathcal{T}_\alpha^n(1), \mathcal{T}_\alpha^n(2), \ldots, \mathcal{T}_\alpha^n(M)$$

   What proportion of intervals cover the target quantity $\tau_\circ$? Try both $n = 100$ and $n = 1000$. Use $M = 10^4$ or larger.

2

Consider data $\mathbf{z} \sim p(\mathbf{z})$ in $\mathcal{Z} = \mathbb{R}^d$ and a Gaussian data model

$$p_\theta(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \mathbf{m}, v\mathbf{I}_d), \quad \boldsymbol{\theta} = \begin{bmatrix} \mathbf{m} \\ v \end{bmatrix} \quad \Theta = \mathbb{R} \times \mathbb{R}_{++}$$

Show that the surprisal loss is

$$\ell_\theta(\mathbf{z}) = \frac{d}{2} \ln v + \frac{1}{2v} \|\mathbf{z} - \mathbf{m}\|^2$$

after omitting constants and that the target parameters are given by:

$$\begin{aligned} \Theta_\circ &= \arg\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) \\ &= \left\{ \begin{bmatrix} \mathbb{E}_\circ[\mathbf{z}] \\ \mathrm{tr}\{\mathbb{V}_\circ[\mathbf{z}]\}/d \end{bmatrix} \right\} \end{aligned}$$

The target parameter is point-identifiable $\Theta_\circ = \{\boldsymbol{\theta}_\circ\}$ in this problem.

What can we say about $\Theta_\circ$ when we instead assume a different Gaussian model

$$p_\theta(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \mathbf{m} + \boldsymbol{\delta}, v\mathbf{I}_d), \quad \boldsymbol{\theta} = \begin{bmatrix} \mathbf{m} \\ \boldsymbol{\delta} \\ v \end{bmatrix} ?$$

3

Consider the our experimental data with detector counts in two settings

$$\mathbf{z} = \begin{bmatrix} c \\ s \end{bmatrix}$$

We want to infer the effective signal rate in an experiment but need to calibrate for background noise.

Given the setting $s = \{0, 1\}$, we use the parametric model:

$$p_\theta(c|s) = \begin{cases} \text{Poisson}(c; \lambda_0), & s = 0 \\ \text{Poisson}(c; \lambda_1), & s = 1 \end{cases} \quad \boldsymbol{\theta} = \begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} \quad \Theta = \mathbb{R}^2_{++}$$

a) Construct the surprisal loss of the (conditional) data model, i.e.,

$$-(1-s) \ln p_\theta(c|s=0) - s \ln p_\theta(c|s=1)$$

Show that after omitting terms that do not depend on $\boldsymbol{\theta}$, it we obtain

$$\ell_\theta(\mathbf{z}) = -(1-s)[c \ln \lambda_0 - \lambda_0] - s[c \ln \lambda_1 - \lambda_1]$$

and that the best parameters

$$\boldsymbol{\theta}_\circ = \begin{bmatrix} \mathbb{E}[c|s=0] \\ \mathbb{E}[c|s=1] \end{bmatrix}$$

This also gives the ERM $\widehat{\boldsymbol{\theta}}_n$ immediately in closed-form.

Tip: Use $\mathbb{E}[\ell_\theta(\mathbf{z})] = \mathbb{E}[\ell_\theta(c, s)] = \mathbb{E}\left[\; \mathbb{E}[\, \ell_\theta(c, s) \mid s \,] \; \right]$

b) Our target quantity is the effective signal rate in the model

$$\tau_\circ = \lambda_{\circ,1} - \lambda_{\circ,0}$$

Generate synthetic training data in the following manner: $n = 2m$ data points. For setting $s = 0$, draw m detection counts c from Poisson$(c; 100)$ and for $s = 1$ another m draws from Poisson$(c; 120)$ Construct the ERM-based confidence intervals

$$\mathcal{T}_\alpha^n = [\tau_{\min}, \tau_{\max}]$$

when $n = 200$. Report the intervals at the 90%, 95% and 99%-confidence levels.

4

Let us return to the ship localization task but now with round-trip time measurements $\mathbf{z}$ using three anchor nodes at locations

$$\mathbf{a}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \mathbf{a}_2 = \begin{bmatrix} 350 \\ 50 \end{bmatrix} \quad \mathbf{a}_3 = \begin{bmatrix} 250 \\ 350 \end{bmatrix}$$

a) We consider a Gaussian data model

$$p_\theta(\mathbf{z}) = \mathcal{N}\Big(\mathbf{z}; 2c^{-1} \underbrace{\begin{bmatrix} \|\mathbf{s} - \mathbf{a}_1\| \\ \|\mathbf{s} - \mathbf{a}_2\| \\ \|\mathbf{s} - \mathbf{a}_3\| \end{bmatrix}}_{\boldsymbol{\mu}(\mathbf{s})}, v\mathbf{I}_3\Big) \quad \boldsymbol{\theta} = \begin{bmatrix} \mathbf{s} \\ v \end{bmatrix}$$

where $c = 3 \times 10^8$. Given n training data vectors $\mathbf{z}^n$ set up the empirical risk $L(\boldsymbol{\theta}, p_n)$ using surprisal loss.

Show that after minimizing with respect to the variance parameter v, we end up with

$$\widehat{\mathbf{s}}_n = \arg\min_{\mathbf{s}} \mathbb{E}_n \left[\|\mathbf{z} - 2c^{-1}\boldsymbol{\mu}(\mathbf{s})\|^2 \right]$$

as our estimated ship location parameter.

b) Generate synthetic training data $\mathbf{z}^n$ in well-specified case where $p(\mathbf{z}) = p_{\theta_\circ}(\mathbf{z})$ and

$$\mathbf{s}_\circ = \begin{bmatrix} 200 \\ 200 \end{bmatrix} \qquad v_\circ = (10^{-7})^2$$

For $n = 100$, plot the contours of the cost function $V(\mathbf{s}) = \mathbb{E}_n \left[\|\mathbf{z} - 2c^{-1}\boldsymbol{\mu}(\mathbf{s})\|^2 \right]$ and compute the ERM $\widehat{\mathbf{s}}_n$ using a gradient search technique as in HW1.