SLDM Homework 3

Dave Zachariah

February 13, 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. Students are expected to attempt each problem.

1

We observe i.i.d. data $\mathbf{z}^n$ for the number of earthquakes each year.

We consider a Poisson model class

$$p_\theta(z) = \text{Poisson}(z; \theta) \qquad \theta > 0$$

and want to validate a learned model.

a) Generate training data $\mathbf{z}^n$ from

$$p(z) = \text{Poisson}(z; 40)$$

Then use the model $p_\theta(z)$ to generate K synthetic datasets $\widetilde{\mathbf{z}}^n_{(1)}, \dots, \widetilde{\mathbf{z}}^n_{(K)}$ to numerically evaluate the probability

$$\begin{aligned} P(\theta) &= \mathbb{P}_\theta\left(T_\theta(\widetilde{\mathbf{z}}^n) \leq T_\theta(\mathbf{z}^n)\right) \\ &\simeq \frac{1}{K}\sum_{k=1}^{K} 1\left\{T_\theta(\widetilde{\mathbf{z}}^n_{(k)}) \leq T_\theta(\mathbf{z}^n)\right\} \end{aligned} \tag{1}$$

Report α_θ for parameters $\theta = 30$ and 40, respectively. Consider $n = 5$, 50 and 500, respectively, and use a large K.

b) Use ERM $\widehat{\theta}_n$ on training data $\mathbf{z}^n$ and report α_θ for $\theta = \widehat{\theta}_n$ ($n = 5$, 50 and 500, respectively).

c) Repeat b) for but now consider data

$$p(z) = \text{NegBinomial}(z; r, p) \qquad r \in \mathbb{N}, p = \frac{40}{r+40}$$

Consider two cases: $r = 10$ and $r = 10^4$.

2

Consider the ship localization problem from HW2 using timing data and the Gaussian data model

$$p_\theta(\mathbf{z}) = \prod_{m=1}^{M} \mathcal{N}\big(z_m; 2c^{-1}\|\mathbf{s} - \mathbf{a}_m\|, v\big) \qquad \boldsymbol{\theta} = \begin{bmatrix} \mathbf{s} \\ v \end{bmatrix}$$

where $c = 3 \times 10^8$ and $M = 3$.

Generate $\mathbf{z}^n$ i.i.d. samples from the following distributions and report α_θ for $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_n$ for $n = 10$, 100 and 1000.

a) Gaussian data:

$$p(\mathbf{z}) = \prod_{m=1}^{M} \mathcal{N}\big(z_m; \mu_m, (10^{-7})^2\big)$$

b) Exponential data:

$$p(\mathbf{z}) = \prod_{m=1}^{M} \mathrm{Exp}\big(z_m; \mu_m\big)$$

where $\mu_m = 2c^{-1}\|[200\ 200]^\top - \mathbf{a}_m\|$

3

We'll continue the localization task from HW2 but consider the small sample case, writing the Gaussian data model as

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

Where we consider a learning parameter based on the maximum posterior belief (aka. MAP):

$$\widehat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}} b(\boldsymbol{\theta}|\mathbf{z}^n)$$

using a prior belief distribution for the parameters:

$$b(\boldsymbol{\theta}) = \mathcal{N}(\mathbf{s}; \mathbf{s}_0, v_0\mathbf{I})$$

so that we consider uniform beliefs for the noise variance $v > 0$.

a) Use the results from the lecture slides to show that we obtain a regularized learning method:

$$\widehat{\boldsymbol{\theta}}_n \equiv \arg\min_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta}, p_n) + \frac{1}{2v_0 n}\|\mathbf{s} - \mathbf{s}_0\|^2$$

and note happens when $v_0 \to \infty$ and what this means in the prior belief distribution. Also note that for any fixed n, the variance parameter v_0 can be chosen to offset any amount of data.

Hint: use the logarithm of $b(\boldsymbol{\theta}|\mathbf{z}^n)$.

b) Generate synthetic training data from $p(\mathbf{z})$ in HW2. We specify the prior belief $b(\boldsymbol{\theta})$ by

$$\mathbf{s}_0 = \begin{bmatrix} 50 \\ 50 \end{bmatrix} \qquad v_0 = 10^2$$

Implement the regularized learning method $\widehat{\boldsymbol{\theta}}_n$ above using either grid search or gradient-based search and plot results along with ship location learn using ERM (HW2) when $n = 1$, 10 and 100. Comment on pros and cons using the regularized method.

4

We are now interested in a rather extreme case of $n = 1$, where we observe blood pressure readings from d patients:

$$\mathbf{z} = [z_1, z_2, \ldots, z_d]^\top$$

drawn from $p(\mathbf{z})$. We are interested in reporting a point estimate of the blood pressure for each patient $k = 1, \ldots, d$, but we only obtain a single reading from each patient, so that $n = 1$.

The sensor errors are specified to be ± 20 [mmHg] (95%).

a) First we consider a data model

$$p_\theta(\mathbf{z}) = \prod_{k=1}^{d} \mathcal{N}(z_k; \theta_k, \sigma^2) \qquad \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_d \end{bmatrix},$$

where $\sigma = (20/2) = 10$ according to the sensor error specification.

Using the surprisal loss, show that the target parameter is

$$\boldsymbol{\theta}_\circ = \mathbb{E}[\mathbf{z}]$$

so that each patient blood pressure target is

$$\tau(\boldsymbol{\theta}_\circ) = \mathbf{e}_k^\top \boldsymbol{\theta}_\circ \qquad k = 1, \ldots, d$$

and conclude that

$$\tau(\widehat{\boldsymbol{\theta}}_n) = z_k \qquad k = 1, \ldots, d$$

when using ERM.

b) Next, consider a much simpler model

$$p_\theta(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \mu\mathbf{1}, (v + \sigma^2)\mathbf{I}), \quad \boldsymbol{\theta} = \begin{bmatrix} \mu \\ v \end{bmatrix}$$

which models the variability of blood pressures across patients by an unknown variance v.

Show that ERM with surprisal loss is

$$\widehat{\boldsymbol{\theta}} = \begin{bmatrix} \widehat{\mu} \\ \widehat{v} \end{bmatrix}$$

is given by

$$\widehat{\mu} = \frac{1}{d}\mathbf{1}^\top \mathbf{z} \qquad \widehat{v} = \max\left(0, \frac{1}{d}\|\mathbf{z} - \widehat{\mu}\mathbf{1}\|^2 - \sigma^2\right)$$

c) Consider now a *random* target parameter

$$\tau(z_k; \boldsymbol{\theta}_\circ) = \frac{v_\circ}{v_\circ + \sigma^2} z_k + \frac{\sigma^2}{v_\circ + \sigma^2}\mu_\circ \qquad k = 1, \dots, d$$

Comment on what this means when the sensor errors $\sigma \to 0$ and $\sigma \to \infty$, respectively. How does this differ from the fixed targets $\tau(\boldsymbol{\theta}_\circ)$ considered above?

d) Generate training data $\mathbf{z}$ from $d = 100$ patients

$$p(\mathbf{z}) = \prod_{k=1}^{d} \mathcal{N}(z_k; \mu_k, \sigma^2) \quad \mu_k \sim \mathcal{U}[100, 170] \quad \sigma^2 = 100$$

Plot two different estimates, $\tau(\widehat{\boldsymbol{\theta}}_n)$ and $\tau(z_k; \widehat{\boldsymbol{\theta}}_n)$, respectively, versus μ_k for all $k = 1, \dots, d$. Comment on your results.