3.4 b Parameter estimation (Density estimation) with Laplace prior.

X is a dataset of observables $\vec{x}_n$: $X = (\vec{x}_1, x_2 \dots \vec{x}_N)$

We model them as Gaussians (i.i.d): $x_i \sim \mathcal{N}(\mu, \sigma) \perp x_j$ $(\forall j \neq i)$,

where μ and σ have to be estimated.

From knowing about the problem, we have a prior on μ:

$$\mathbb{P}(\mu) \sim \text{Laplace}(0, b) = \frac{1}{2b} e^{-(|\mu|/b)}$$

The MAP estimate is:

$$\mu_{MAP} = \operatorname{argmax}_\mu \mathbb{P}(\mu, \sigma \mid X = \{\dots\})$$

$$= \text{———} \log\left(\frac{\mathbb{P}(X = \{\dots\} \mid \mu) \; \mathbb{P}(\mu)}{\mathbb{P}(X = \{\dots\})}\right)$$

(We can insert log because it's a monotonously increasing function, and is well defined since probas are > 0).

(Also $\mathbb{P}(X = \{\dots\})$ does not depend on μ, it is a constant w.r.t. the argmax)

$$\mu_{MAP} = \operatorname{argmax}_\mu \left(\sum_{n=1}^{N} \log(\mathbb{P}(X_n = x_n \mid \mu)) + \log \mathbb{P}(\mu) \right)$$

$$= \text{———}_\mu \left(\sum_{n=1} \log\left(\frac{1}{\underbrace{\sqrt{2\pi}\sigma}_{\perp \text{ from } \mu}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}\right) + \log\left(\underbrace{\frac{1}{2b}}_{\perp \text{ from } \mu} e^{-|\mu|/b}\right)\right)$$

$$= \operatorname{argmax}_\mu \left(\sum_{n=1}^{N} -\frac{(x_n-\mu)^2}{2\sigma^2} - \frac{|\mu|}{2b} \right)$$

$$= \underline{\underline{\operatorname{argmin}}}_\mu \left(+ \sum_{n=1}^{N=1} (x_n - \mu)^2 + \frac{\sigma^2}{b} |\mu| \right)$$

we switch the sign and multiply by $2\sigma^2$

We can search for minima by asking the (pseudo) gradient of this to be 0.

$$\mathcal{L} = \sum_{n=1}^{N} (x_n - \mu)^2 + \frac{\sigma^2}{b} |\mu|$$

There are 2 cases:

$\mu \neq 0 \implies \nabla_\mu |\mu| = \pm 1 = \text{sign}(\mu)$

$\mu = 0 \implies \nabla_\mu |\mu| \in [-1, 1]$

(By continuity. One can think of a function like abs(·), but rounded at zero, over a range of ε, with $\varepsilon \to 0$)

$\hookrightarrow$ (To be discussed another time).

<u>Case 1</u>: $\mu_{MAP} \neq 0$.

$$\nabla_\mu \mathcal{L} = 0 \iff \nabla_\mu \sum_n (x_n - \mu)^2 + \frac{\sigma^2}{b} \text{sign}\, \mu = 0$$

$$\iff -2 \sum_{n=1}^{N} (x_n - \mu) + \frac{\sigma^2}{b} \text{sign}\, \mu = 0$$

$$2N\mu - 2\sum x_n + \frac{\sigma^2}{b} \text{sign}(\mu) = 0$$

$$\mu - \bar{x} + \frac{\sigma^2}{2bN} \text{sign}(\mu) = 0 \quad , \bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

2 Sub-cases: $\mu_{MAP} > 0$ and $\mu_{MAP} < 0$

* $\mu > 0$: $\mu - \bar{x} + \frac{\sigma^2}{2bN} \times (+1) = 0$

$$\implies \mu_{MAP} = \bar{x} - \frac{\sigma^2}{2bN} \quad (> 0) \iff \left(\bar{x} > \frac{\sigma^2}{bN}\right)$$

* $\mu < 0$ $\quad \mu - \bar{x} + \frac{\sigma^2}{2bN} \times (-1) = 0$

$$\implies \mu_{MAP} = \bar{x} + \frac{\sigma^2}{2bN} \quad (< 0) \iff \left(\bar{x} < -\frac{\sigma^2}{bN}\right.$$

Remark: in both cases, we notice that $|\bar{x}| > \frac{\sigma^2}{2bN}$.

$, \text{sign}(\mu_{MAP}) = \text{sign}\, \bar{x}$

Case 2 : $\mu = 0$

$$\nabla_\mu \mathcal{L} = 0 \iff -2 \sum_{n=1}^{N} (x_n - \underbrace{\mu}_{0}) + \frac{\sigma^2}{b} \nabla_\mu |\mu| \Big|_{\mu=0} = 0 \quad \Big\downarrow \tfrac{1}{2N}$$

$$-\bar{x} + \frac{\sigma^2}{2bN} \cdot \nabla_\mu |\mu| = 0$$

$$\nabla_\mu |\mu| = \frac{1}{\left(\frac{\sigma^2}{2bN}\right)} \cdot \bar{x}$$

Is this $(\nabla_\mu |\mu|)$ in the range $[-1, 1]$?

It is, iff: $-1 < \nabla_\mu |\mu| < 1$

$$\iff -1 < \frac{1}{\left(\frac{\sigma^2}{2bN}\right)} \cdot \bar{x} < 1$$

$$\iff -\frac{\sigma^2}{2bN} < \bar{x} < \frac{\sigma^2}{2bN}$$

This is complementary to the range found in case 1 ($\mu \neq 0$)

Summary: $\mathcal{L} = \sum_n (x_n - \mu)^2 + \frac{\sigma^2}{b} |\mu|$ is minimized by $\mu = \mu_{MAP}$:

$$\mu_{MAP} = \begin{cases} \text{if } |\bar{x}| > \frac{\sigma^2}{2bN}, & \mu_{MAP} = \bar{x} - \text{sign}(\bar{x}) \cdot \frac{\sigma^2}{2bN} \\ \text{if } |\bar{x}| < \frac{\sigma^2}{2bN}, & \mu_{MAP} = 0. \end{cases}$$

Is this sol° continuous in $\bar{x}$?

When $\bar{x} \to \left(\frac{\sigma^2}{2bN}\right)^+ \quad \left(= \frac{\sigma^2}{2bN} + \varepsilon, \ \varepsilon \to 0^+\right)$,

$$\mu_{MAP} = \bar{x} - \frac{\sigma^2}{2bN} = \varepsilon = 0^+$$

Yes, the sol° is $\mathcal{C}^0$: