Probset #2

3. (d)

Density of Laplace distribution

$f_{\mathcal{L}}(z \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|z-\mu|}{b}\right)$

linear regression model $y = x^T\theta + \epsilon \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$

Laplace prior on this model $\theta \sim \mathcal{L}(0, bI)$

Show that θ_{MAP} in this case is equivalent to the solution of linear regression with L1 regularization, whose loss is specified as

$J(\theta) = \| X\theta - \vec{y} \|_2^2 + r\|\theta\|_1$

Also, what is the value for r?

To optimize this, use gradient descent with a random initialization and solve it numerically.

prior = $\theta \sim \mathcal{L}(0, bI)$

$p(\theta) = f_{\mathcal{L}}(\theta \mid 0, bI) = \frac{1}{2b} \exp\left(-\frac{|\theta|}{b}\right)$

$$
\begin{aligned}
\theta_{MAP} &= \underset{\theta}{\operatorname{argmax}}\ p(y \mid \theta, x)\, p(\theta) \\
&= \underset{\theta}{\operatorname{argmax}}\ p(y \mid \theta, x) \times \frac{1}{2b} \exp\left(-\frac{|\theta|}{b}\right) \\
&= \underset{\theta}{\operatorname{argmin}} - \log\left\{ p(y \mid \theta, x) \times \frac{1}{2b} \exp\left(-\frac{|\theta|}{b}\right) \right\} \\
&= \underset{\theta}{\operatorname{argmin}} \left[-\log p(y \mid \theta, x) - \log \frac{1}{2b} + \frac{|\theta|}{b} \right] \\
&= \underset{\theta}{\operatorname{argmin}} \left[-\log p(y \mid \theta, x) + \frac{|\theta|}{b} \right]
\end{aligned}
$$

$$= \underset{\theta}{\operatorname{argmin}} \left[-\log p(y|\theta, x) + \frac{|\theta|}{b} \right]$$

$$= \underset{\theta}{\operatorname{argmin}} \left\{ -\log \prod_{i=1}^{m} \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2} \left(y^{(i)} - \theta^T x^{(i)}\right)^2 \right\} \right] + \frac{|\theta|}{b} \right\}$$

$$= \underset{\theta}{\operatorname{argmin}} \left\{ -\log \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^m + \sum_{i=1}^{m} \frac{1}{2\sigma^2} \left(y^{(i)} - \theta^T x^{(i)}\right)^2 + \frac{|\theta|}{b} \right\}$$

$$= \underset{\theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{m} \frac{1}{2\sigma^2} \left(y^{(i)} - \theta^T x^{(i)}\right)^2 + \frac{|\theta|}{b} \right\}$$

<use gradient descent>

$$l(\theta) = \frac{1}{m} \left\{ \sum_{i=1}^{m} \frac{1}{2\sigma^2} \left(\underset{(1,1)}{y^{(i)}} - \underset{(1,n)}{\theta^T} \underset{(n,1)}{x^{(i)}}\right)^2 + \frac{|\theta|}{b} \right\}$$

$$\underset{\theta}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma^2} \sum_{i=1}^{m} \|X\theta - \vec{y}\|_2^2 + \frac{|\theta|}{b} \right\}$$

$$\underset{\theta}{\operatorname{argmin}} \frac{1}{2\sigma^2} \left\{ \sum_{i=1}^{m} \|X\theta - \vec{y}\|_2^2 + \frac{2\sigma^2}{b} \times |\theta| \right\}$$

$$\gamma = \frac{2\sigma^2}{b}$$

* Remark

- Linear Regression w/ L_2 regularization : Ridge regression
- Linear Regression w/ L_1 regularization : Lasso regression

  ↳ Known to result in sparse parameters, where most of the parameter values are zero, with some of them non-zero.