CS229 Problem Set 2
Solution to Problem 3

<Parth Bansal>

June 16, 2025

Problem 3

(a)

We start from Bayes' rule:

$$p(\theta \mid x, y) = \frac{p(x, y, \theta)}{p(x, y)} = \frac{p(y \mid x, \theta)p(x, \theta)}{p(x, y)} = \frac{p(y \mid x, \theta)p(\theta \mid x)p(x)}{p(x, y)}$$

Assume $p(\theta) = p(\theta \mid x)$, then:

$$p(\theta \mid x, y) = \frac{p(y \mid x, \theta)p(\theta)p(x)}{p(x, y)} = p(y \mid x, \theta)p(\theta) \cdot \frac{p(x)}{p(x, y)}$$

So:

$$\theta_{\text{MAP}} = \arg\max_\theta p(\theta \mid x, y) = \arg\max_\theta p(y \mid x, \theta)p(\theta) \cdot \frac{p(x)}{p(x, y)} = \arg\max_\theta p(y \mid x, \theta)p(\theta)$$

(b)

$\theta_{\text{MAP}} = \arg\max_\theta p(y \mid x, \theta)p(\theta) = \arg\max_\theta \log(p(y \mid x, \theta)p(\theta)) = \arg\max_\theta \log p(y \mid x, \theta) + \log p(\theta)$

This is equivalent to:

$$\theta_{\text{MAP}} = \arg\min_\theta - \log p(y \mid x, \theta) - \log p(\theta)$$

Assume:

$$\theta \sim \mathcal{N}(0, \eta^2 I)$$

Then:

$$p(\theta) = \frac{1}{(2\pi)^{n/2}\eta^n} \exp\left\{-\frac{1}{2\eta^2}\theta^T\theta\right\} = (2\pi)^{-n/2}\eta^{-n} \exp\left\{-\frac{1}{2\eta^2}\|\theta\|_2^2\right\}$$

Taking the logarithm:

$$\log p(\theta) = -\frac{n}{2}\log(2\pi) - n\log\eta - \frac{1}{2\eta^2}\|\theta\|_2^2$$

Then:

$$\theta_{\text{MAP}} = \arg\min_\theta - \log p(y \mid x, \theta) + \frac{1}{2\eta^2}\|\theta\|_2^2$$

Define:

$$\lambda = \frac{1}{2\eta^2}$$

(c)

$$\begin{aligned}
\epsilon^{(i)} &\sim \mathcal{N}(0, \sigma^2) \\
y^{(i)} &= \theta^T x^{(i)} + \epsilon^{(i)} \\
y^{(i)} \mid x^{(i)}, \theta &\sim \mathcal{N}(\theta^T x^{(i)}, \sigma^2) \\
p(y^{(i)} \mid x^{(i)}, \theta) &= \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{1}{2\sigma^2}(y^{(i)} - \theta^T x^{(i)})^2\right\} \\
p(\vec{y} \mid X, \theta) &= \prod_{i=1}^{m} p(y^{(i)} \mid x^{(i)}, \theta) \\
&= \frac{1}{(2\pi)^{m/2}\sigma^m} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(y^{(i)} - \theta^T x^{(i)})^2\right\} \\
&= \frac{1}{(2\pi)^{m/2}\sigma^m} \exp\left\{-\frac{1}{2\sigma^2}\|X\theta - \vec{y}\|_2^2\right\}
\end{aligned}$$

$$\begin{aligned}
\log p(\vec{y} \mid X, \theta) &= -\frac{m}{2}\log(2\pi) - m\log\sigma - \frac{1}{2\sigma^2}\|X\theta - \vec{y}\|_2^2 \\
\theta_{\text{MAP}} &= \arg\min_{\theta} -\log p(\vec{y} \mid x, \theta) + \frac{1}{2\eta^2}\|\theta\|_2^2 \\
&= \arg\min_{\theta} \frac{1}{2\sigma^2}\|X\theta - \vec{y}\|_2^2 + \frac{1}{2\eta^2}\|\theta\|_2^2 \\
J(\theta) &= \frac{1}{2\sigma^2}\|X\theta - \vec{y}\|_2^2 + \frac{1}{2\eta^2}\|\theta\|_2^2 \\
\nabla_\theta J(\theta) &= \frac{1}{\sigma^2}(X^T X\theta - X^T\vec{y}) + \frac{1}{\eta^2}\theta = 0 \\
\theta_{\text{MAP}} &= (X^T X + \frac{\sigma^2}{\eta^2} I)^{-1} X^T \vec{y}
\end{aligned}$$

(d)

$$\begin{aligned}
\theta &\sim \mathcal{L}(0, bI) \\
p(\theta) &= \frac{1}{(2b)^n} \exp\left\{-\frac{1}{b}\|\theta\|_1\right\} \\
\log p(\theta) &= -n\log(2b) - \frac{1}{b}\|\theta\|_1 \\
\theta_{\text{MAP}} &= \arg\min_{\theta} \frac{1}{2\sigma^2}\|X\theta - \vec{y}\|_2^2 - \log p(\theta) \\
&= \arg\min_{\theta} \frac{1}{2\sigma^2}\|X\theta - \vec{y}\|_2^2 + \frac{1}{b}\|\theta\|_1 \\
J(\theta) &= \|X\theta - \vec{y}\|_2^2 + \gamma\|\theta\|_1 \\
\theta_{\text{MAP}} &= \arg\min_{\theta} J(\theta) \\
\gamma &= \frac{2\sigma^2}{b}
\end{aligned}$$