

$$6a) p(\theta|x,y) = \frac{p(\theta,xy)}{p(xy)} \stackrel{\text{def}}{=} p(y|x,\theta) \frac{p(\theta|y)}{p(y|x)} = p(y|x,\theta) \frac{p(\theta)p(x)}{p(y|x)p(x)}$$

$p(y|x)$ indep. of θ .

$$\downarrow \hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}} \, p(y|x,\theta) \frac{p(\theta)}{p(y|x)} = \underset{\theta}{\operatorname{argmax}} \, p(y|x,\theta)p(\theta)$$

$$6b) \text{ If we set our prior to be } p(\theta|\mathcal{O}, \mu^2 I) = \frac{1}{\sqrt{2\pi^k}} \exp\left(-\frac{1}{2\mu^2}\theta^T\theta\right)$$

$$\text{then we get that } \hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}} \, p(y|x,\theta) \sim \exp(-\cdot)$$

$$\begin{aligned} \hat{\theta}_{MAP} &= \underset{\theta}{\operatorname{argmax}} \log p(y|x,\theta) - \log(p(y|x,\theta)) - \frac{\theta^T\theta}{2\mu^2} \\ &= \underset{\theta}{\operatorname{argmin}} -\log p(y|x,\theta) + (\frac{1}{2\mu^2})\|\theta\|^2, \quad \lambda = \frac{1}{2\mu^2} \end{aligned}$$

$$6c) \text{ In other words, we have } p(\vec{y}|X,\theta) \sim \exp\left(-\frac{1}{2\sigma^2}(\vec{y}-X\theta)^T(\vec{y}-X\theta)\right)$$

$$\Rightarrow \hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmin}} \frac{1}{2\sigma^2} (\vec{y}-X\theta)^T(\vec{y}-X\theta) + \frac{\theta^T\theta}{2\mu^2} = \underset{\theta}{\operatorname{argmin}} \frac{1}{2\sigma^2} \left[\vec{y}^T\vec{y} + \theta^T X^T X \theta - 2\theta^T X^T \vec{y} + \frac{1}{2\mu^2} \theta^T \theta \right]$$

Taking the derivative w.r.t θ of the loss gives :

$$\frac{\partial}{\partial \theta} \text{Loss} = \frac{1}{2} (x^T x \Theta - 2x^T y) + \frac{\Theta}{\sigma^2} = 0$$

$$\Leftrightarrow \cancel{(x^T x \Theta - 2x^T y)} + (\frac{\sigma^2}{2} x^T x + \frac{\lambda}{2} \Theta^T \Theta) \Theta_{MAP} = x^T y \cancel{\Theta}$$

$$\Rightarrow \cancel{\Theta^T \Theta} = \cancel{x^T x \Theta - 2x^T y} + \cancel{\Theta^T x^T y} \quad \Theta_{MAP} = \underline{\left[x^T x + \frac{\sigma^2}{2} \mathbb{I} \right]^{-1}} \underline{x^T y}$$

$$= \underline{\left(x^T x + 2\sigma^2 \lambda \mathbb{I} \right) x^T y}$$

6d) Now $\underline{\Theta_{MAP}} = \underset{\Theta}{\operatorname{argmin}} -\log p(y|x,\Theta) + \sum_{i=1}^k \frac{1}{b} |\Theta_i|_1$

$$= \underset{\Theta}{\operatorname{argmin}} \frac{1}{2\sigma^2} \|\vec{y} - X\vec{\Theta}\|_2^2 + \frac{1}{b} |\Theta|_1$$

$$= \underset{\Theta}{\operatorname{argmin}} \|\vec{y} - X\vec{\Theta}\|_2^2 + \frac{\alpha\sigma^2}{b} |\Theta|_1, \text{ with } \gamma = \frac{2\sigma^2}{b}$$