Maximum A Posteriori Estimation (MAP)

$$f_{Bayes}(\vec{x}) = \mathbb{E}[Y \mid \vec{X} = \vec{x}]$$

$$= \int_y y \, p_{Y|\vec{X}}(y \mid \vec{x}) \, dy$$

If $p_{Y|\vec{X}}$ is a function of some parameter $w \in \mathcal{W}$ then we just need to estimate that parameter

$$\text{MLE: } \underset{w \in \mathcal{W}}{\arg\max} \, \underbrace{p(D \mid w)}_{} = \prod_{i=1}^{n} p(z_i \mid w)$$

"find w that maximizes the likelihood of the data"

$$\text{MAP: } \underset{w \in \mathcal{W}}{\arg\max} \, \underbrace{p(w \mid D)}_{\text{"posterior"}}$$

"find w that is the most likely given the data"

MAP Basics

Ex: $Z_i \sim \mathcal{N}(\mu^*, 1)$ is the i-th persons height

$$p_z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z-\mu^*)^2}{2}\right)$$

$\mu \sim \mathcal{N}(160, \sigma^2)$, $\quad p_\mu(\mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\mu-160)^2}{2\sigma^2}\right)$

$$= \underset{\mu \in \mathbb{R}}{\arg\min} \left[\sum_{i=1}^{n} \frac{(z_i - \mu)^2}{2} - \log\left(\frac{1}{\sqrt{2\pi}}\right) + \frac{(\mu - 160)^2}{2\sigma^2} \right]$$

$$= \underset{\mu \in \mathbb{R}}{\arg\min} \left[\sum_{i=1}^{n} \frac{(z_i - \mu)^2}{2} + \frac{(\mu - 160)^2}{2\sigma^2} \right]$$

$$\frac{dg}{d\mu}(\mu) = \sum_{i=1}^{n} (z_i - \mu) + \frac{(\mu - 160)}{\sigma^2} = 0$$

$$\Rightarrow \sum_{i=1}^{n} z_i - n\mu + \frac{\mu}{\sigma^2} - \frac{160}{\sigma^2} = 0$$

$$\Rightarrow \sum_{i=1}^{n} z_i - \frac{160}{\sigma^2} = \left(n - \frac{1}{\sigma^2}\right)\mu$$

$$\Rightarrow \mu = \frac{\sum_{i=1}^{n} z_1 - \frac{160}{\sigma^2}}{n - \frac{1}{\sigma^2}} = \mu_{MAP}$$

Estimating $\vec{w}$ for $P_{Y|\vec{X}}$

$$D = ((\vec{X}_1, Y_1), \ldots, (\vec{X}_n, Y_n)) \in (\mathcal{X} \times \mathcal{Y})^n, \ \mathbb{P}_D, \ p_D$$

$(\vec{X}_i, Y_i)$ are i.i.d with $\mathbb{P}_{\vec{X},Y}$ and $p_{\vec{X},Y}$

Assume $Y_i | \vec{X}_i = \vec{x}_i \sim \mathcal{N}(\vec{x}_i^T \vec{w}^*, 1)$

$$p_{Y|\vec{X}=\vec{x}}(y|\vec{x}) = \frac{1}{\sqrt{2\pi}} \exp\left(- \frac{(y - \vec{x}^T \vec{w}^*)^2}{2} \right)$$

Calculating $\vec{w}_{MAP}$:

$$\vec{w}_{MAP} = \underset{\vec{w} \in \mathbb{R}^{d+1}}{\operatorname{argmax}} \; p(\vec{w} \mid D)$$

$$= \underset{\vec{w} \in \mathbb{R}^{d+1}}{\operatorname{argmax}} \; p(D \mid \vec{w})\, p(\vec{w})$$

$$= \underset{\vec{w} \in \mathbb{R}^{d+1}}{\operatorname{argmin}} \; -\log\left(p(D \mid \vec{w})\, p(\vec{w})\right)$$

$$= \underset{\vec{w} \in \mathbb{R}^{d+1}}{\operatorname{argmin}} \left[-\log\left(p(D \mid \vec{w})\right) - \log\left(p(\vec{w})\right)\right]$$

$$= \underset{\vec{w} \in \mathbb{R}^{d+1}}{\operatorname{argmin}} \left[\sum_{i=1}^{n} \frac{(y_i - \vec{x}_i^T \vec{w})^2}{2} - \log\left(p(\vec{w})\right)\right]$$

$$= \underset{\vec{w} \in \mathbb{R}^{d+1}}{\arg\min} \left[\sum_{i=1}^{n} \frac{(y_i - \vec{x}_i^T \vec{w})^2}{2} - \log(p(w_0)) - d \log\left(\frac{1}{\sqrt{2\pi/\lambda}} \right) + \sum_{j=1}^{d} \frac{\lambda w_j^2}{2} \right]$$

$$= \underset{\vec{w} \in \mathbb{R}^{d+1}}{\arg\min} \left[\sum_{i=1}^{n} \frac{(y_i - \vec{x}_i^T \vec{w})^2}{2} + \frac{\lambda}{2} \sum_{j=1}^{d} w_j^2 \right]$$

$$= \underset{\vec{w} \in \mathbb{R}^{d+1}}{\arg\min} \ \frac{n}{2} \hat{L}_\lambda$$

$$f_{Bayes}(\vec{x}) = \mathbb{E}[Y \mid \vec{X} = \vec{x}]$$

$$= \int_y y \, p_{Y|\vec{X}}(y \mid \vec{x}) \, dy$$

$$\approx \int_y y \, p(y \mid \vec{x}, \vec{w}_{MAP}) \, dy$$

$$= \mathbb{E}[Y' \mid \vec{X} = \vec{x}] \quad \text{where } Y' \mid \vec{X} = \vec{x} \sim \mathcal{N}(\vec{x}^T \vec{w}_{MAP}, 1)$$

$$= \vec{x}^T \vec{w}_{MAP}$$

$$= \vec{x}^T \hat{\vec{w}}_\lambda$$

$$= \hat{f}$$

Assume $W_j \sim \text{Laplace}(0, 1/\lambda)$ are i.i.d for $j \in \{1, \dots, d\}$

and $W_0 \sim \text{Laplace}(0, b)$ for very large b

$\simeq \text{Uniform}(-a, a)$ for large a

$$\vec{w}_{MAP} = \underset{\vec{w} \in \mathbb{R}^{d+1}}{\arg\max}\; p(\vec{w} \mid D)$$

$$= \underset{\vec{w} \in \mathbb{R}^{d+1}}{\arg\min} \left[\sum_{i=1}^{n} \frac{(y_i - \vec{x}_i^T \vec{w})^2}{2} + \lambda \sum_{j=1}^{d} |w_j| \right]$$

Lasso regression