Maximum A Posteriori Estimation (MAP)

want to estimate $p_{Y|\vec{X}}$

$$\vec{w}_{MLE} = \underset{\vec{w} \in \mathcal{W}}{\operatorname{argmax}} \; \overbrace{p_D(D|\vec{w})}^{\text{Likelihood}}$$

$$\vec{w}_{MAP} = \underset{\vec{w} \in \mathcal{W}}{\operatorname{argmax}} \; \overbrace{p(\vec{w}|D)}^{\text{Posterior}}$$

Bayes rule

$$= \underset{\vec{w} \in \mathcal{W}}{\operatorname{argmax}} \; \frac{p(\vec{w}, D)}{p(D)}$$

chain rule

$$= \underset{\vec{w} \in \mathcal{W}}{\operatorname{argmax}} \; \frac{p(D|\vec{w})\,p(\vec{w})}{p(D)}$$

$$= \underset{\vec{w} \in \mathcal{W}}{\operatorname{argmax}} \; \underbrace{p(D|\vec{w})}_{\text{likelihood}}\,\underbrace{p(\vec{w})}_{\text{prior}}$$

MAP Basics

Ex: $z_i \overset{i.i.d.}{\sim} \mathcal{N}(\mu^*, 1)$ height of i-th person

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

Assume

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

$$\mu_{MAP} = \underset{\mu \in \mathbb{R}}{\operatorname{argmax}}\ p(D|\mu)\,p(\mu)$$

$$= \underset{\mu \in \mathbb{R}}{\operatorname{argmax}} \prod_{i=1}^{n} p(z_i|\mu)\ p(\mu)$$

$$= \underset{\mu \in \mathbb{R}}{\operatorname{argmin}} -\sum_{i=1}^{n} \log\Big(p(z_i|\mu)\ p(\mu)\Big)$$

$$= \underset{\mu \in \mathbb{R}}{\operatorname{argmin}} -\sum_{i=1}^{n} \Big[\log\big(p(z_i|\mu)\big) + \log\big(p(\mu)\big)\Big]$$

$$= \underset{\mu \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \left[\frac{(z_i-\mu)^2}{2}\right] - \log\big(p(\mu)\big)$$

$$\log(p(\mu)) = \log\left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\mu-160)^2}{2\sigma^2}\right)\right)$$

$$= \log\left(\frac{1}{\sqrt{2\pi}}\right) + \log\left(\exp\left(-\frac{(\mu-160)^2}{2\sigma^2}\right)\right)$$

$$= \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{(\mu-160)^2}{2\sigma^2}$$

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

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

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

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

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

$$\Rightarrow \mu_{MAP} = \frac{\sum_{i=1}^{n} z_i + \frac{160}{\sigma^2}}{n + 1/\sigma^2} \qquad \mu_{MLE} = \frac{\sum_{i=1}^{n} z_i}{n}$$

if $\frac{1}{\sigma^2} \gg \sum_{i=1}^{n} z_i$ and $\frac{1}{\sigma^2} \gg n$ (i.e., σ^2 is small)

then $\mu_{MAP} \approx \frac{\frac{160}{\sigma^2}}{1/\sigma^2} = 160$

if $\frac{1}{\sigma^2} \ll \sum_{i=1}^{n} z_i$ and $\frac{1}{\sigma^2} \ll n$ (i.e., σ^2 is large)

then $\mu_{MAP} \approx \frac{\sum_{i=1}^{n} z_i}{n} = \mu_{MLE}$

Estimating $p_{Y|\vec{X}}$

$$D = \left((\vec{X}_1, Y_1), \ldots, (\vec{X}_n, Y_n)\right) \qquad \mathbb{P}_D, \; p_D$$

$$(\vec{X}_i, Y_i) \overset{i.i.d.}{\sim} \mathbb{P}_{\vec{X},Y}, \quad p_{\vec{X},Y}$$

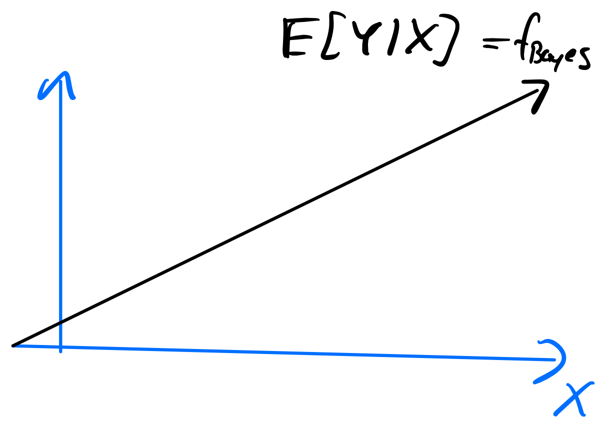
chain rule

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

Assume:

$$Y|\vec{X}=\vec{x} \sim \mathcal{N}(\vec{x}^T\vec{w}^*, 1) \qquad \vec{x}^T\vec{w}^* = \mathbb{E}[Y|\vec{X}=\vec{x}]$$

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



Assume

$$w_j \sim \mathcal{N}(0, 1/\lambda)$$

i.i.d. for $j \in \{1, \ldots, d\}$

$w_0 \sim \mathcal{N}(0, \sigma^2)$ for a very large σ^2

$\approx$ Uniform$(-a, a)$ with a very large a

and independent of w_j for $j \in \{1, \ldots, d\}$

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

$$= \underset{\vec{w} \in \mathbb{R}}{\operatorname{argmax}} \; \prod_{i=1}^{n} p(\vec{x}_i, y_i \mid \vec{w}) \;\; p(\vec{w})$$

$$= \underset{\vec{w} \in \mathbb{R}}{\operatorname{argmin}} \; -\sum_{i=1}^{n} \log\left(p(\vec{x}_i, y_i \mid \vec{w}) \;\; p(\vec{w})\right)$$

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

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

$$\log\left(p(\vec{w})\right) = \log\left(p(w_0, w_1, \ldots, w_d)\right)$$

$$= \log\left(p(w_0)\, p(w_1) \cdots p(w_d)\right)$$

$$= \log\left(p(w_0) \prod_{j=1}^{d} p(w_j)\right)$$

$$= \log\left(p(w_0)\right) + \log\left(\prod_{j=1}^{d} p(w_j)\right)$$

$$= \log(p(w_0)) + \sum_{j=1}^{d} \log(p(w_j))$$

$$= \log(p(w_0)) + \sum_{j=1}^{d} \log\left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(w_j)^2}{2\,1/\lambda}\right)\right)$$

$$= \log(p(w_0)) + d\log\left(\frac{1}{\sqrt{2\pi}}\right) - \lambda \sum_{j=1}^{d} \frac{w_j^2}{2}$$

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

$$= \operatorname*{argmin}_{\vec{w} \in R} \underbrace{\frac{1}{n} \sum_{i=1}^{n} \left[\frac{(\vec{x}_i^T \vec{w} - y_i)^2}{2}\right.}_{\hat{L}(\vec{w})} + \underbrace{\left.\frac{\lambda}{n} \sum_{j=1}^{d} \frac{w_j^2}{2}\right]}_{\text{squared regularizer}}$$

$$= \operatorname*{argmin}_{\vec{w} \in R} \hat{L}_\lambda(\vec{w})$$

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

$$= \int_{\mathcal{Y}} y \, P_{Y|\vec{X}=\vec{x}}(y \mid \vec{x}, \vec{w}^*) \, dy$$

$$\approx \int_{\mathcal{Y}} y \, P_{Y|\vec{X}=\vec{x}}(y \mid \vec{x}, \vec{w}_{MAP}) \, dy$$

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

$$= \vec{x}^T \vec{w}_{MAP} = \hat{f}_\lambda(\vec{x})$$

$$\hat{f}_\lambda = \arg\min_{f \in \mathcal{F}_1} \hat{L}_\lambda(f)$$

Lasso Regression

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

$w_0 \sim \text{Laplace}(0, b)$ for large b

and independent for $j \in \{1, \ldots, d\}$

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

$$p(w_j) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{|w_j|}{2} \right)$$