Linear Regression – Probabilistic Version

Model: $y_n = w^T x_n$

Prob. view: $y_n = w^T x + \varepsilon_n$
$\varepsilon_n \sim \mathcal{N}(0, \sigma^2)$

i.e.: $P(y, x \mid w) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^D \cdot e^{-\frac{1}{2\sigma^2}(y - wx)^2}$

MLE: $P(X = (x,y) \mid w) = \prod_n^N P(y_n, x_n \mid w)$

$\frac{\partial \mathcal{L}}{\partial w} = 0 = \frac{\partial}{\partial w} \log \prod_n^N \left(\frac{1}{\sigma^2}\right)^{D/2} \cdot e^{-\!-\!-}$

$= \frac{\partial}{\partial w} \sum_n^N -\frac{1}{2\sigma^2} \cdot (y - wx)^2 + \underbrace{\frac{\partial}{\partial w} \log \left(\frac{1}{\sigma^2}\right)^{D/2}}_{0}$

$0 = \sum_n^N (y - wx)^2 \iff \text{LSE}.$

MAP: $P(w) \sim \mathcal{N}(0, \tau^2)$

$\mathcal{L} = P(w \mid X) = \frac{P(X \mid w) P(w)}{P(X)}$

$\arg\max_w \mathcal{L} = \arg\max_w \left(\log P(X \mid w) + \log P(w)\right)$

$\frac{\partial \mathcal{L}}{\partial w} = 0 = \frac{\partial}{\partial w}\left(-\frac{1}{2\sigma^2} \sum_n^N (y - wx)^2\right) + \frac{\partial}{\partial w} \log \left(\underbrace{\frac{1}{\tau^2}}_{0} e^{-\frac{w^2}{2\tau^2}}\right)$

$= \frac{\partial}{\partial w}\left(\quad\right) - \frac{\partial}{\partial w} \frac{w^2}{2\tau^2}$

$= \frac{\partial}{\partial w}\left(-\frac{1}{2\sigma^2} \sum_n^N (y - wx)^2 - \frac{w^2}{2\tau^2}\right)$

$\iff \text{LSE} + \underbrace{\frac{\sigma^2}{\tau^2}}_{\lambda} \|w\|^2$

λ ! Ridge!