regularized linear regression

TC

October 2021

I.

$$p(w \mid x, t, \alpha, \beta) \propto p(t \mid x, w, \beta) p(w \mid \alpha)$$

$$=> log(p(w \mid x, t, \alpha, \beta) \propto log(p(t \mid x, w, \beta) p(w \mid \alpha))$$

we have:

$$p(t \mid x, w, \beta) = \prod_{n=1}^{N} N(t_n \mid y(x_n, w), \beta^{-1})$$

$$p(t \mid x, w, \beta) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\beta^{-1}}} \times e^{\dfrac{-(t - y(x_n, w))^2}{2\beta^{-2}}}$$

$$log(p(t \mid x, w, \beta)) = \frac{-\beta}{2} \Sigma_{n=1}^{N} (t - y(x_n, w))^2 + noise$$

And:

$$p(\mathbf{w} \mid \alpha) = N(w \mid 0, \alpha^{-1} I) = \frac{1}{(2\pi)^{D/2} \mid \Sigma \mid^{1/2}} e^{\dfrac{-(w-0)^T \Sigma^{-1} (w-0)}{2}}$$

$$log(p(\mathbf{w} \mid \alpha)) = \frac{-1}{2} w^T w + noise$$

So:

$$log(p(w \mid x, t, \alpha, \beta) \propto \frac{-\beta}{2} \Sigma_{n=1}^{N} (t - y(x_n, w))^2 + \frac{-1}{2} w^T w$$

we find that the maximum of the posterior is given by the minimum of:

$$\frac{\beta}{2} \Sigma_{n=1}^{N} (t - y(x_n, w))^2 + \frac{1}{2} w^T w$$

or we minimize:

$$Q = \mid\mid X\mathbf{w} - \mathbf{t} \mid\mid_2^2 + \lambda \mathbf{w}^\mathbf{T} \mathbf{w}$$

$$\bigtriangledown Q_w = 2X^T (X\mathbf{w} - \mathbf{t}) + 2\lambda \mathbf{w}$$
$$\rightarrow \mathbf{w} = (X^T X + \lambda I)^{-1} X^T \mathbf{t}$$