Bayesian Linear Regression

$$p(y|\Phi, w) = \mathcal{N}(y|\Phi w + b, \Sigma)$$

$$p(w) = \mathcal{N}(w|m_0, S_0)$$

say we want the MAP estimate

1. get posterior
2. take $\text{gradient}_w = 0$

Linear Algebra Tricks

$$(AB)^T = B^T A^T$$

$$a^T b = b^T a \quad \text{for vectors}$$

$$\nabla_x (a^T x) = a$$

$$\nabla_x (x^T A x) = (A + A^T)x$$

for simplicity, we take the $-\log$ posterior & minimize it.

$$-\log(p(w|\Phi, y, m_0, S_0, \Sigma, b)) = \frac{1}{2}(y - \Phi w - b)^T \Sigma^{-1} (y - \Phi w - b) + \frac{1}{2}(w - m_0) S_0^{-1} (w - m_0)$$

Simplify (discuss what these simplifications mean)

$$b = 0 \qquad m_0 = 0$$

$$\Sigma^{-1} = \beta I \qquad S_0^{-1} = \lambda I$$

after simplifying

$$-\log p(w|\cdot) = \beta \underbrace{(y-\Phi w)^T(y-\Phi w)}_{\text{expand this}} + \lambda w^T w$$

$$\overset{A}{y^Ty} \overset{B}{- (\Phi w)^T y - y^T \Phi w} + \overset{C}{(\Phi w)^T \Phi w}$$

not gonna matter in ∇_w

$\cancel{B \rightarrow \boxed{-w^T\Phi^T} y - \boxed{(\Phi w)^T y} \rightarrow -2}$ vectors!

vectors!

$$B \quad -\boxed{y^T\Phi w} - y^T \Phi w$$

$$= -2(\Phi^T y)^T w$$

$$C \quad w^T \Phi^T \Phi w$$

$$\nabla_w -\log p(w|\cdot) = \beta(-2\Phi^T y + 2\Phi^T\Phi w) + 2\lambda w$$

set $= 0$

$\Downarrow$

$$\hat{w}^{MAP} = (\beta \Phi^T \Phi + \lambda I)^{-1} \beta \Phi^T y$$

this is the mean of the posterior gaussian

3

Discuss

Basis Functions Global vs. Local

Demo App: How does basis functions meet up with hyperplane fitting?

Discriminative Vs Generative models

Use ~~$p(y|\Phi w+b)$~~ $p(y|\Phi, w)$ in a classifier

$$p(\text{class}, y, \Phi) = p(\text{class})\, p(y, \Phi | \text{class})$$

$$p(y, \Phi | \text{class}) = p(\Phi)\, p(y | \Phi, w_{\text{class}})$$