

PENALTY — FOR OVERLY COMPLICATED (ABLE TO OVERFIT)

E.G. AKAIKE-CRITERION:

MIN (ERROR + NUMBER OF PARAMETERS)

SOME FUNCTION
E.G. $LM(M)$

PENALTY — FOR LARGE WEIGHT VALUES — REGULARIZATION

(BISHOP 3.11.4)

$$E(\underline{w}) = \left[\frac{1}{2} \sum_{n=1}^N \{y(x_n, \underline{w}) - t_n\}^2 + \frac{\lambda}{2} \|\underline{w}\|^2 \right]$$

$\frac{\partial}{\partial \underline{w}}$ ↓ (CHECK IT!)

$$\underline{w}_{LS} = (\underline{\Phi}^T \underline{\Phi} + \lambda \underline{I})^{-1} \underline{\Phi}^T \underline{t}$$

$$\sum_{k=1}^M w_k^2$$

BEST VALUE FOR λ ? — MORE CHOICES

↓
MORE MODELS

↓
VALIDATION → (NEW) VALIDATION DATA SET

REGULARIZATION & PARAMETER NORM

$$E(\underline{w}) = \text{ERROR} + \frac{\lambda}{2} \|\underline{w}\|^q$$

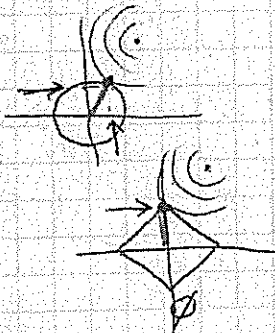
$$\sum_{i=1}^n |w_i|^q$$

$q = 2$

EASY COMPUTATION

$q = 1$

SPARSE COEFFICIENTS
(NON DIFFERENTIABLE!)

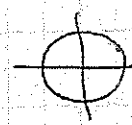


$q = 1/2$

(BISHOP FIG 3.3)



$q = 1$



$q = 2$



$q = 4$



$q = \infty$