

- DIFFERENT ACTIVATION FUNCTIONS

- COMPUTATIONAL EFFICIENCY

- FAST COMPUTATION OF JACOBIAN $J_{ac} = \frac{\partial y_h}{\partial x_i}$

& HESSIAN (INVERSE, PRODUCTS, ETC.)

$$\frac{\partial^2 E}{\partial w_{ji} \partial w_{ik}}$$

↳ BETTER GRADIENT METHODS

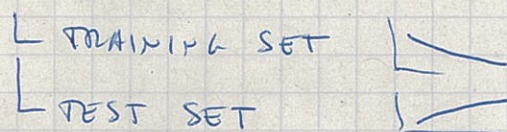
REGULARIZATION (CH. 5.5, FIG. 5.9)

$$\tilde{E}(\underline{w}) = E(\underline{w}) + \frac{\lambda}{2} \underline{w}^T \underline{w} \quad \text{WEIGHT DECAY}$$

↳ PROBLEMS WITH SCALING

$$p(\underline{w}) \propto \exp\left(-\frac{\lambda}{2} \sum_u \alpha_u \|\underline{w}\|_u^2\right)$$

EARLY STOPPING (CH. 5.5.2)



(BEST GENERALIZATION)

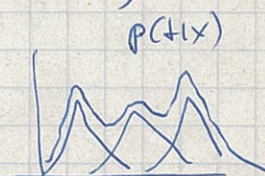
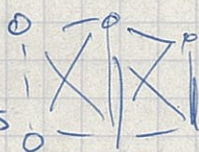
$$\sum_{j \in w_k} w_j^2$$

↳ ANY DIVISION INTO GROUPS

CONVOLUTIONAL NETWORKS (DEEP LEARNING) (CH. 5.5.6)

MIXTURE DENSITY NETWORKS

OUTPUTS: - MIXING COEFFICIENTS
- KERNEL WIDTHS
- KERNEL CENTRES



(FIG. 5.20)

(INVERSE PROBLEMS)

BAYESIAN NEURAL NETWORKS

- PRIOR DISTRIBUTION ON WEIGHTS

- LEARNING POSTERIOR FROM DATA (LAPLACE-APP., ETC.)

- INFERENCE ON FUTURE DATA

(AISPAC.ORG/NEURAL/)