

μ is on p. 12-21

EE 434
02/26/07

$$H = J^T J \quad (J \text{ in (12.23), Jacobian})$$

can be singular

used in eq. (12.28) for weight update in training

$$x_{k+1} = x_k - [J^T(x_k) J(x_k)]^{-1} J^T(x_k) v(x_k) \quad (12.28)$$

$$H = J^T J \approx \text{Hessian}$$

as H can be singular add a term with a parameter

$$N \rightarrow H \rightarrow G = H + \mu I \quad I = \text{identity}$$

to choose a μ ; find eigenvalues of G which will be those of $H + \mu I$;
(12.30)

$$G z_i = (H + \mu I) z_i = \lambda_i z_i = H z_i + \mu z_i \\ = \underbrace{(\text{eigenvalue of } H)}_{\lambda_i} z_i + \mu z_i$$

$$= (\mu + \lambda_i) z_i$$

can choose $\mu > \max_i |\lambda_i|$ then $\mu + \lambda_i > 0$ for all i

and G becomes nonsingular.

To get (12.28) they use the error function (12.20)

$$F(x) = \sum_{i=1}^N v_i^2(x) = v^T(x) \cdot v(x); \quad N = \begin{matrix} M & Q \\ \parallel & \parallel \\ \text{\# of} & \text{\# of training} \\ \text{output} & \text{examples} \\ \text{neurons} & \end{matrix}$$

use Newton's method result (12.19)

$$x_{k+1} = x_k - A_k^{-1} g_k$$

x = all weights & biases

see p. 9.11 (9.43) and p. 8.13 (8.38)

$$(8.38) \text{ assumes } F(x) = \frac{1}{2} x^T A x + d^T x + c \quad (8.35)$$

$$\frac{\partial F}{\partial x_i} \Rightarrow \nabla F : \quad = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j + \sum_{k=1}^n d_k x_k + c$$

$$\frac{\partial F}{\partial x_l} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \frac{\partial x_i x_j}{\partial x_l} + \sum_{k=1}^m d_k \frac{\partial x_k}{\partial x_l} + 0$$

(for $l=1, \dots, m$
 $x = m$ -vector)

$$= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \left[\underbrace{x_i}_{1 \text{ if } i=l} \frac{\partial x_j}{\partial x_l} + \underbrace{x_j}_{1 \text{ if } j=l} \frac{\partial x_i}{\partial x_l} \right] + d_l \quad \begin{matrix} 1 \text{ if } k=l \\ 0 \text{ if } k \neq l \end{matrix}$$

$$a_{ij} = a_{ji}$$

$$\text{or } A = A^T$$

is symmetric

$$= \frac{1}{2} \sum_{i=1}^n [a_{il} x_i] + \frac{1}{2} \sum_{j=1}^n a_{lj} x_j + d_l$$

$$= \frac{1}{2} \left(\sum_{i=1}^n [a_{li} x_i + \overbrace{a_{li} x_i}^{2a_{li} x_i}] \right) + d_l$$

$\uparrow$
 j is a dummy index

$$\Rightarrow \nabla F = \begin{bmatrix} \partial F / \partial x_1 \\ \vdots \\ \partial F / \partial x_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n a_{1i} x_i \\ \vdots \\ \sum_{i=1}^n a_{mi} x_i \end{bmatrix} + \begin{bmatrix} d_1 \\ \vdots \\ d_m \end{bmatrix} = Ax + d \quad (8.38)$$

$\underbrace{\nabla^2 F(x)}_{=A}$

$$F = \frac{1}{2} x^T A x + d x + c$$

next $\Rightarrow$ p. 9-11, (9.43), Δx

$$\text{Eq. (9.40)} \quad F(x + \Delta x) = F(x) + \nabla F \cdot \Delta x + \frac{1}{2} \Delta x^T A \Delta x$$

Eq. (9.41) desire a minimum

$$\nabla F = 0 \Rightarrow A \Delta x + d \stackrel{\nabla F \text{ at } x}{=} \Rightarrow \text{set } 0 = A \Delta x + \nabla F \quad (9.41)$$

$$\Rightarrow \Delta x = -A^{-1} \nabla F(x) \quad (9.42)$$

desire to update x to $x + \Delta x$; needs A^{-1} to explicit

LM algorithm uses this in (12.19) top of p. 12.20

$$x(k+1) = x(k) - A^{-1}(k) \nabla F(k) \quad (12.19)$$

see m-file arma1.m