ECE544 NA

10/15/2013

TODAY

NEWTON

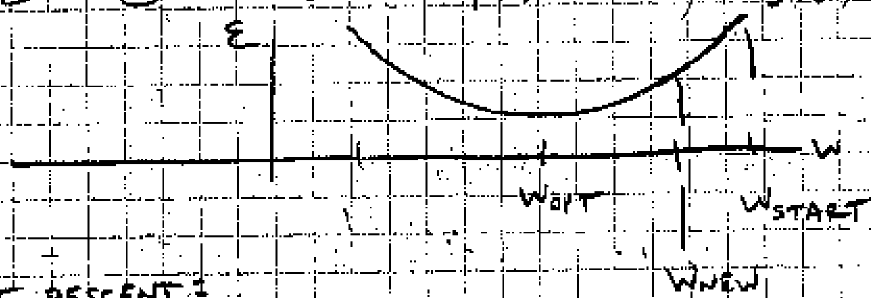
LEVENBERG-MARQUARDT

MATRIX INVERSION

~~[illegible]~~

NEWTON OPTIMIZATION

SUPPOSE $\mathcal{E} \approx a(w - w_{opt})^2 + b$, a, b, w_{opt} UNKNOWN



GRADIENT DESCENT:

$$w_{NEW} = w_{START} - \eta \frac{\partial \mathcal{E}}{\partial w}$$

MIGHT BE TOO FAR, OR NOT FAR ENOUGH

IN TERMS OF UNKNOWN a, b:

$$\frac{\partial \mathcal{E}}{\partial w} = 2a(w - w_{opt})$$

$$\frac{\partial^2 \mathcal{E}}{\partial w^2} = 2a$$

NEWTON ITERATION:

$$\boxed{w_{NEW} = w_{START} - \frac{\partial \mathcal{E}/\partial w}{\partial^2 \mathcal{E}/\partial w^2}}$$

IF $\mathcal{E}$ TRULY QUADRATIC, THEN $w_{NEW} = w_{opt}$!

MULTI-DIMENSIONAL

$$\vec{W}_{NEW} = \vec{W}_{START} - H^{-1} \nabla \mathcal{E}$$

$$\nabla \mathcal{E} = \underbrace{\begin{bmatrix} \partial \mathcal{E}/\partial w_1 \\ \vdots \\ \partial \mathcal{E}/\partial w_W \end{bmatrix}}_{[W \times 1]}, \quad H = \underbrace{\begin{bmatrix} \frac{\partial^2 \mathcal{E}}{\partial w_1^2} & \frac{\partial^2 \mathcal{E}}{\partial w_1 \partial w_2} & & \\ & & \ddots & \\ & & & \frac{\partial^2 \mathcal{E}}{\partial w_W^2} \end{bmatrix}}_{[W \times W]}$$

REMEMBER

① ERROR BACKPROP FINDS ∇E
IN $O\{\overline{W}\}$
- ALREADY A BIG NUMBER, E.G.
$\overline{W} = (D+1)M + (M+1)C$ OR
$\overline{W} = \frac{1}{2}M(M+1)$

② H HAS $\overline{W}^2$ ELEMENTS!
INVERTING H IS $O\{\overline{W}^3\}$!!

· LEVENBERG-MARQUARDT : FIND H AS A SUM
OF RANK-1 MATRICES
(RANK-1 : AN $O\{\overline{W}\}$ REPRESENTATION
THOUGH COMPUTING H IS $O\{\overline{W}^2\}$)
MATRIX INVERSION LEMMA : FIND H^{-1} IN
$O\{\overline{W}^2\}$

LEVENBERG-MARQUARDT :

ASSUME $E = \frac{1}{2}\sum_n (y^n - t^n)^2$

Then $\frac{\partial^2 E}{\partial w_{ji}\partial w_{lk}} = \sum_n \frac{\partial y^n}{\partial w_{ji}}\frac{\partial y^n}{\partial w_{lk}} + \sum_n (y^n - t^n)\frac{\partial^2 y^n}{\partial w_{ji}\partial w_{lk}}$

DIGRESSION : MMSE ESTIMATORS

SUPPOSE $y(x) = \arg\min E_{t}\left[(y(x) - t)^2 \mid x\right]$

$= \arg\min\left[y^2(x) - 2y(x)E[t|x] + E[t^2|x]\right]$

DIFFERENTIATE :

$2y(x) - 2E[t|x] = 0$

SO $\boxed{y(x) = \arg\min E[(y(x)-t)^2] \Leftrightarrow y(x) = E[t|x]}$

LIKEWISE

$$y(x) = \arg\min \sum_n \left(y(x^n) - t^n\right)^2$$

$$\Leftrightarrow \sum_n \left(y(x^n) - t^n\right) = 0 \; !$$

IF $\dfrac{\partial^2 y^n}{\partial w_{ji} \partial w_{lk}}$ UNCORRELATED WITH y^n

THEN $$\sum_n \left(y^n - t^n\right) \frac{\partial^2 y^n}{\partial w_{ji} \partial w_{lk}} = 0$$

AND

$$H = \frac{\partial^2 E}{\partial w_{lk} \partial w_{ji}} = \sum_n \frac{\partial E^n}{\partial w_{ji}} \frac{\partial E^n}{\partial w_{lk}} = \sum_n \delta_j^n z_i^n \delta_l^n z_k^n$$

RANK ONE MATRIX

REMEMBER

$$\nabla E = \frac{\partial E}{\partial w_{ji}} = \sum_n \frac{\partial E^n}{\partial w_{ji}} = \sum_n \delta_j^n z_i^n$$

CALL $\vec{g}^n = \begin{bmatrix} \vdots \\ \frac{\partial E^n}{\partial w_{ji}} \\ \vdots \end{bmatrix}$.

THEN $H = \sum_n \vec{g}^n \vec{g}^{nT}$

BUT WE WANT H^{-1}!

MATRIX INVERSION LEMMA

$$(A + BC)^{-1} = A^{-1} - A^{-1}B\left(I + CA^{-1}B\right)^{-1}CA^{-1}$$

ACCUMULATE

$$H^N = \sum_{n=1}^{N} \vec{g}^n \vec{g}^{nT}$$

$$H^{N+1} = H^N + \vec{g}^{N+1}\left(\vec{g}^{N+1}\right)^T$$

$$\left(H^{N+1}\right)^{-1} = \left(H^N + \vec{g}^{N+1}\left(\vec{g}^{N+1}\right)^T\right)^{-1}$$

$$= \left(H^N\right)^{-1} - \left(H^N\right)^{-1}\vec{g}^{N+1}\underbrace{\left(I + \overbrace{\left(\vec{g}^{N+1}\right)^T\left(H^N\right)^{-1}\vec{g}^{N+1}}^{O\{W\}}\right)^{-1}}_{\text{SCALAR}}\underbrace{\left(\vec{g}^{N+1}\right)^T\left(H^N\right)^{-1}}_{O\{W^2\}}$$