ARTIFICIAL NEURAL NETWORK

Swetha R.

Department of Electronics and Communication Engineering

UNCONSTRAINED OPTIMIZATION TECHNIQUE

Class 7:

Newton's Method

Swetha R.

Department of Electronics and Communication Engineering

OUTLINE

- Unconstrained optimization technique
  - Newton's Method
  - Gauss-Newton method

Swetha R.
Department of Electronics and Communication Engineering

- The idea is to minimize the quadratic approximation of the cost function E (w) around the current point w(n).
- This minimization is performed at each iteration, using 2nd order Taylor series expansion of E(w) around the point w(n)
  - $E(w(n) + \Delta w(n)) = E\big(w(n)\big) + E'(w(n))\Delta w(n) + \frac{1}{2}\Delta w^T(n)E''\big(w(n)\big)\Delta w(n) + \dots$

- Note: Taylor series expansion

$$f(\mathbf{x}_k + \Delta\mathbf{x}) \approx f(\mathbf{x}_k) + \mathbf{g}_k^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T Q_k \Delta\mathbf{x}_k$$

,where

$$\mathbf{g}_k = \mathbf{g}(\mathbf{x}_k) = \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_k}$$

- Gradient of cost function i.e., diff. w.r.to $\Delta w(n)$
  - $\nabla E(w(n) + \Delta w(n)) = E'\big(w(n)\big) + \frac{1}{2}.2.H\big(w(n)\big)\Delta w(n) = 0$
  - $E'\big(w(n)\big) = g(n)$ =1st derivative

$$\mathbf{H} = \nabla^2\mathscr{E}(\mathbf{w})$$

$$= \begin{bmatrix} \dfrac{\partial^2\mathscr{E}}{\partial w_1^2} & \dfrac{\partial^2\mathscr{E}}{\partial w_1 \partial w_2} & \cdots & \dfrac{\partial^2\mathscr{E}}{\partial w_1 \partial w_m} \\ \dfrac{\partial^2\mathscr{E}}{\partial w_2 \partial w_1} & \dfrac{\partial^2\mathscr{E}}{\partial w_2^2} & \cdots & \dfrac{\partial^2\mathscr{E}}{\partial w_2 \partial w_m} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial^2\mathscr{E}}{\partial w_m \partial w_1} & \dfrac{\partial^2\mathscr{E}}{\partial w_m \partial w_2} & \cdots & \dfrac{\partial^2\mathscr{E}}{\partial w_m^2} \end{bmatrix}$$

- Gradient of cost function i.e., diff. w.r.to $\Delta w(n)$

  - $\nabla E(w(n)+\Delta w(n)) = E'\big(w(n)\big) + \frac{1}{2}.2.H\big(w(n)\big)\Delta w(n) = 0$

$\Rightarrow H\big(w(n)\big)\Delta w(n) =- E'\big(w(n)\big)$

$\Rightarrow \Delta w(n) = \ -H\big(w(n)\big)^{-1}E'\big(w(n)\big)$

$\Rightarrow w(n+1) = w(n) - \eta H\big(w(n)\big)^{-1}E'\big(w(n)\big)$

$\Rightarrow w(n+1) = w(n) - \eta H\big(w(n)\big)^{-1} E'\big(w(n)\big)$

$\Rightarrow$ Where $H^{-1} = inverse\ of\ Hessian\ matrix$

$\Rightarrow$ For Newton's method to work well H must be positive definite,

Newton's Method Example:

For example, Let us consider X(i) and D(i) are from stochastic process(ergodic), then error signal will also be stochastic.

Cost Function ,$\xi \triangleq E[e^2(k)]$, E - Expectation

In matrix form, $e(k) = D - W^T X$

$$e^2(k) = (D - W^T X)(D - W^T X)^T$$
$$= DD^T - DX^T W - W^T X D^T + W^T X X^T W$$

$$\Rightarrow\ \xi = E[e^2(k)] = E[DD^T - DX^TW - W^TXD^T +\ W^TXX^TW\ ]$$

$$= r_d - r_{dx}W - W^Tr_{xd} + W^TR_xW$$

Where, $r_d = E[DD^T]$, $R_x = E[XX^T]$

$r_{dx} = E[DX^T] = r_{xd} = E[XD^T]$

UNCONSTRAINED OPTIMIZATION TECHNIQUE

Newton's Method

$$\Rightarrow \frac{\partial \xi}{\partial W} = -r_{dx} - r_{xd} + W^T R_x + R_x W$$

$$= -2r_{dx} + 2R_x W = E'\big(w(n)\big)$$

$$\frac{\partial^2 \xi}{\partial W^2} = 2R_x = H\big(w(n)\big)$$

UNCONSTRAINED OPTIMIZATION TECHNIQUE

Newton's Method

$$w(n+1) = w(n) - \eta H\big(w(n)\big)^{-1} E'\big(w(n)\big)$$

$$w(n+1) = w(n) - \eta \frac{1}{2} {R_x}^{-1} (-2r_{dx} + 2R_x W)$$

$$w(n+1) = w(n) - \eta\big(\mathrm{w(n)} - w_{op}\big)$$

$$= (1-\eta)w(n) + \eta w_{op}$$

Newton's Method

With affine transformation and rotation, the above equation is transformed in the form of x(k) = $\alpha^k x(0)$, which is stable and converge to zero if $|\alpha| < 1$

With similar analysis, $|1-\eta| < 1 \Rightarrow 0 < \eta < 2$

UNCONSTRAINED OPTIMIZATION TECHNIQUE

Newton's Method:

Generally, *Newton's* method *converges quickly* and *does not exhibit the zigzagging behavior* of the method of *steepest descent*.

However, *Newton's* method has *two main disadvantages*:

- the *Hessian* matrix H(n) has to be a *positive definite* matrix for all n, which is *not guaranteed* by the algorithm.
    - This is solved by the *modified Gauss-Newton* method.
- It has high *computational complexity*

UNCONSTRAINED OPTIMIZATION TECHNIQUE

Gauss-Newton Method:

Applicable to the cost function of sum of error squares, Let

$$E(w) = \frac{1}{2}\sum_{i=1}^{n} e^2(i)$$

e(i) is function of w, which is fixed over observation interval $1 \leq i \leq n$

Gauss-Newton Method:

e(i) is function of w, which is fixed over observation interval $1 \leq i \leq n$.

Given an operating point w(n), linearize the dependence of e(i) on w

as

$$e'(i,w) = e(i) + \left[\frac{\partial e(i)}{\partial w}\right]^T \big(w - w(n)\big),\ i = 1,2,\ldots,n \ \ at\, w = w(n)$$

Gauss-Newton Method:

In matrix form

$$e'(n,w) = e(n) + J(n)\big(w - w(n)\big)$$

where, e(n) is the error vector , e(n) =$[e(1), e(2), \ldots, e(n)]^T$ and J(n) is the

Jacobian matrix, which is defined as

$$\mathbf{J}(n) = \begin{bmatrix} \frac{\partial e(1)}{\partial w_1} & \frac{\partial e(1)}{\partial w_2} & \cdots & \frac{\partial e(1)}{\partial w_m} \\ \frac{\partial e(2)}{\partial w_1} & \frac{\partial e(2)}{\partial w_2} & \cdots & \frac{\partial e(2)}{\partial w_m} \\ \vdots & \vdots & & \vdots \\ \frac{\partial e(n)}{\partial w_1} & \frac{\partial e(n)}{\partial w_2} & \cdots & \frac{\partial e(n)}{\partial w_m} \end{bmatrix}_{\mathbf{w}=\mathbf{w}(n)}$$

Gauss-Newton Method:

The gradient of error vector

$$\nabla e(n) = [\nabla e(1), \nabla e(2), \ldots, \nabla e(n)] \text{ m x n}$$

But, J(n) is the Jacobian matrix of dimension n x m, therefore

J(n) = $\nabla e(n)^T$

The updated weight w(n+1) is defined by

$$w(n+1) = arg\min_{w}\left\{\frac{1}{2}\|e'(n,w)\|^2\right\}$$

UNCONSTRAINED OPTIMIZATION TECHNIQUE

Gauss-Newton Method:

The squared Euclidean norm of error vector

$$\frac{1}{2}\|e'(n,w)\|^2$$

$$= \frac{1}{2}\|e\,(n)\|^2 + e^T(n)J(n)\big(w - w(n)\big) + \frac{1}{2}\big(w - w(n)\big)^T J^T(n)J(n)(w - w(n))$$

Diff. this eq. with w and setting it to zero , we get

UNCONSTRAINED OPTIMIZATION TECHNIQUE

Gauss-Newton Method:

The squared Euclidean norm of error vector

$$\frac{1}{2}\|e'(n,w)\|^2$$

$$= \frac{1}{2}\|e\,(n)\|^2 + e^T(n)J(n)\big(w - w(n)\big) + \frac{1}{2}\big(w - w(n)\big)^T J^T(n)J(n)(w - w(n))$$

Diff. this eq. with w and setting it to zero , we get

$$J^T(n)e(n) + J^T(n)J(n)\big(w - w(n)\big) = 0$$

Method Of Steepest Descent:

Let a, b $\in \mathbb{R}^n$, then

$$\frac{\partial(a^T b)}{\partial a} = \frac{\partial(b^T a)}{\partial a} \triangleq \begin{pmatrix} \frac{\partial(b^T a)}{\partial a_1} \\ \dots \\ \frac{\partial(b^T a)}{\partial a_n} \end{pmatrix} = \begin{pmatrix} b_1 \\ \dots \\ b_n \end{pmatrix} = b$$

And Let P = $P^T \in \mathbb{R}^{nxn}$

$$\frac{\partial(a^T P a)}{\partial a} = 2Pa$$

Gauss-Newton Method:

The squared Euclidean norm of error vector

$$\frac{1}{2}\|e'(n,w)\|^2$$

$$= \frac{1}{2}\|e\,(n)\|^2 + e^T(n)J(n)\big(w - w(n)\big) + \frac{1}{2}\big(w - w(n)\big)^T J^T(n)J(n)(w - w(n))$$

Diff. this eq. with w and setting it to zero , we get

$$J^T(n)e(n) + J^T(n)J(n)\big(w - w(n)\big) = 0$$

Gauss-Newton Method:

$$J^T(n)e(n) + J^T(n)J(n)\big(w - w(n)\big) = 0$$

$$\Rightarrow w - w(n) = -\left(J^T(n)J(n)\right)^{-1} J^T(n)e(n)$$

$$\Rightarrow w(n+1) = w(n) - \left(J^T(n)J(n)\right)^{-1} J^T(n)e(n)$$

Gauss-Newton Method:

Gauss-Newton Method *does not require Hessian matrix as that* of the *Newton's* method instead it uses Jacobian of error matrix

However, Gauss-Newton Method has a *disadvantage*:

- $J^T(n)J(n)$ matrix has to be nonsingular, which is *not guaranteed* always
  - This is solved by adding δI to $J^T(n)J(n)$ matrix
- Where δ is a positive constant, added to ensure
- $J^T(n)J(n) + \delta I$ to be positive definite for all n

UNCONSTRAINED OPTIMIZATION TECHNIQUE

Gauss-Newton Method:

Therefore, the modified weight update equation Gauss-Newton Method as

$$w(n+1) = w(n) - (J^T(n)J(n) + \delta I)^{-1} J^T(n)e(n)$$

THANK YOU

Swetha R.
Department of Electronics and Communication Engineering

swethar@pes.edu

+91 80 2672 1983 Extn 753