Hatırlatma

Geriye Yayılım Algoritması (Back-Propagation Algorithm)

Verilenler: $\left\{\left(x^l, y_d^l\right)\right\}\Big|_{l=1}^{P}$ Eğitim Kümesi

Hesaplananlar: Eğitim Kümesindeki l. çifte ilişkin çıkış katmanındaki j. nörondaki hata $e_j^l = y_{d_j}^l - y_j^l$

Nöron j için ani hata: $\frac{1}{2}(e_j^l)^2$

Toplam ani hata: $E^{(l)} = \frac{1}{2}\sum_{j\in C}(e_j^l)^2$ Neden sadece çıkış katmanı?

Ortalama karesel hata: $E_{ort} = \frac{1}{p}\sum_{l=1}^{p} E^{(l)}$ Ağdaki hangi büyüklüklere bağlı?

Hatırlatma

Verilen eğitim kümesi için, ortalama karesel hata E_{ort}'yı öğrenme performansının ölçütü olarak al ve bu amaç ölçütünü enazlayan parametreleri belirle.

EK BİLGİ

Bazı Eniyileme (Optimizasyon) Teknikleri

Eniyileme problemi

$$\min_{x\in\Omega} f(x) \qquad f(x): R^n \to R$$

Kısıtlar: $h(x)=0 \qquad g(x)\le 0$

$$h(x): R^n \to R^p \qquad g(x): R^n \to R^m$$

Kısıtsız Eniyileme Problemi

$$\min_{x\in R^n} f(x)$$

Hatırlatma

Teorem: $\Omega \in R^n, \quad f(.) \in C^1$

1. Mertebeden gerek koşul $x^* \in \Omega$ $f(.)$'in ekstremum noktası ise

$$\nabla f(x^*) = 0$$

Teorem: $\Omega \in R^n, \quad f(.) \in C^2, \quad x^* \in \Omega$

2. Mertebeden yeter koşul

$(i)\ \nabla f(x^*) = 0$

Nasıl hesaplanır? $(ii)\ H(x^*)$ kesin pozitif

→ $x^* \in \Omega$ $f(.)$'in minimum noktasıdır.

Doğrultu Belirleme (Line Search) Algoritması

- $x^{(0)}$ başlangıç noktasını belirle
- s doğrultusunu belirle
- $f(x^{(k)} + \alpha s^{(k)})$'yı α'ya göre enazlayan $\alpha^{(k)}$'yı belirle
- $x^{(k+1)} = x^{(k)} + \alpha^{(k)} s^{(k)}$ doğrultusunu belirle

Hatırlatma

Amaç: $f(x^{(k+1)}) < f(x^{(k)})$

Beklenti: Algoritma fonksiyonu enazlayan x^*'a yakınsayacak

Ne zaman sona erdilecek? $\left|f(x^{(k+1)}) - f(x^{(k)})\right| < \varepsilon$

- s doğrultusunu belirle Nasıl ?

$* s^{(k)} \hat{=} -\nabla f(x^{(k)})$ → "en dik iniş " (steepest descent)

$* \alpha^{(k)} s^{(k)} \hat{=} -H^{-1}(x^{(k)})\nabla f(x^{(k)})$ → Newton Metodu

$* \alpha^{(k)} s^{(k)} \hat{=} -\frac{1}{2}[J^{(k)^T} J^{(k)}]^{-1}\nabla f(x^{(k)})$ → Gauss-Newton Metodu

Bu doğrultuların işe yaradığını nasıl gösterebiliriz?

Gauss-Newton Metodu

$$\alpha^{(k)} s^{(k)} \triangleq -\frac{1}{2}[J^{(k)^T} J^{(k)}]^{-1} \nabla f(x^{(k)}) \text{ ile } f(x^{(k+1)}) < f(x^{(k)})$$

sağlanır mı?

Kısıtlama: $f(x) \triangleq \frac{1}{2} r(x)^T r(x)$

$$J^{(k)} \triangleq \frac{\partial r(x)^T}{\partial x}\Big|_{x=x(k)} \rightarrow \nabla f(x^{(k)}) = J^{(k)^T} r(x^{(k)})$$

Gauss-Newton Metodu'nun amacı özel bir $f(x)$ için Hessian matrisini kullanmadan 2. mertebe yöntem kadar iyi sonuç elde etmek.

$$x^{(k+1)} = \min_x \{ \frac{1}{2} \| \overbrace{r(x^{(k)}) + J^{(k)}(x - x^{(k)})}^{p(x)} \|^2 \}$$

Bu ifade aslında nedir?

$$\frac{1}{2}\|p(x)\|^2 = \frac{1}{2}\|r(x^{(k)})\|^2 + r(x^{(k)})^T J^{(k)}(x - x^{(k)})$$

$$+ \frac{1}{2}(x - x^{(k)})^T J^{(k)^T} J^{(k)} (x - x^{(k)})$$

$$\frac{d\left(\frac{1}{2}\|p(x)\|^2\right)}{dx} = 0 \rightarrow J^{(k)^T} r(x^{(k)}) + J^{(k)^T} J^{(k)}(x - x^{(k)}) = 0$$

$$x = x^{(k)} - (J^{(k)^T} J^{(k)})^{-1} J^{(k)^T} r(x^{(k)})$$

$$x^{(k+1)} = \min_x \{ \frac{1}{2} \| r(x^{(k)}) + J^{(k)}(x - x^{(k)}) \|^2 \}$$

$$x^{(k+1)} = x^{(k)} - (J^{(k)^T} J^{(k)})^{-1} J^{(k)^T} r(x^{(k)})$$

$(J^{(k)^T} J^{(k)})^{-1}$ varsa $\rightarrow$ $f(x^{(k+1)}) < f(x^{(k)})$

Sonuç: x^*'a yakınsayacak

EK BİLGİNİN SONU

Amaç: Verilen eğitim kümesi için, ortalama karesel hata E_{ort} 'yı öğrenme performansının ölçütü olarak al ve bu amaç ölçütünü enazlayan parametreleri belirle.

Toplam ani hata: $E^{(l)} = \frac{1}{2} \sum_{j \in C} (e_j^l)^2$

Ortalama karesel hata: $E_{ort} = \frac{1}{p} \sum_{l=1}^{p} E^{(l)}$

Yapılan: E_{ort} yerine E 'yi en azlamak

Eğitim kümesindeki *k.* veri için ileri yolda hesaplananı yazalım:

1. Gizli Katman Çıkışı

$$\begin{bmatrix} v_1^{(1)} \\ v_2^{(1)} \\ \vdots \\ v_{ns1}^{(1)} \end{bmatrix} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & \cdots & w_{1(gs+1)}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & \cdots & w_{2(gs+1)}^{(1)} \\ \vdots & \vdots & & \vdots \\ w_{ns11}^{(1)} & w_{ns12}^{(1)} & & w_{ns1(gs+1)}^{(1)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{gs} \\ 1 \end{bmatrix} \xrightarrow{\varphi(.)} \begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \vdots \\ y_{ns1}^{(1)} \end{bmatrix}$$

2. Gizli Katman Çıkışı

$$\begin{bmatrix} v_1^{(2)} \\ v_2^{(2)} \\ \vdots \\ v_{ns2}^{(2)} \end{bmatrix} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & \cdots & w_{1(ns1+1)}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & \cdots & w_{2(ns1+1)}^{(2)} \\ \vdots & \vdots & & \vdots \\ w_{(ns2)1}^{(2)} & w_{(ns2)2}^{(2)} & & w_{ns2(ns1+1)}^{(2)} \end{bmatrix} \begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \vdots \\ y_{ns1}^{(1)} \\ 1 \end{bmatrix} \xrightarrow{\varphi(.)} \begin{bmatrix} y_1^{(2)} \\ y_2^{(2)} \\ \vdots \\ y_{ns2}^{(2)} \end{bmatrix}$$

$$\begin{bmatrix} v_1^{(o)} \\ v_2^{(o)} \\ \vdots \\ v_q^{(o)} \end{bmatrix} = \begin{bmatrix} w_{11}^{(o)} & w_{12}^{(o)} & \cdots & w_{1(nsgks+1)}^{(o)} \\ w_{21}^{(o)} & w_{22}^{(o)} & \cdots & w_{2(nsgks+1)}^{(o)} \\ \vdots & \vdots & & \vdots \\ w_{q1}^{(o)} & w_{q2}^{(o)} & & w_{q(nsgks+1)}^{(o)} \end{bmatrix} \begin{bmatrix} y_1^{(gks)} \\ y_2^{(gks)} \\ \vdots \\ y_{nsgks}^{(gks)} \\ 1 \end{bmatrix} \xrightarrow{\varphi(.)} \begin{bmatrix} y_1^{(o)} \\ y_2^{(o)} \\ \vdots \\ y_q^{(o)} \end{bmatrix} \rightarrow \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_q \end{bmatrix}$$

$$v^{(1)} = w^{(1)} \begin{bmatrix} x \\ 1 \end{bmatrix} \Rightarrow y^{(1)}$$

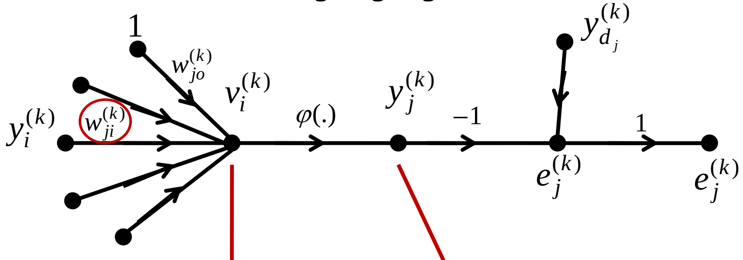
$$v^{(2)} = w^{(2)} \begin{bmatrix} y^{(1)} \\ 1 \end{bmatrix} \Rightarrow y^{(2)}$$

$$\mathrm{E}^{(k)} = \frac{1}{2} \sum_{j \in C} (e_j^{(k)})^2 = \frac{1}{2} e^{(k)^T} e^{(k)}$$

$$v^{(o)} = w^{(o)} \begin{bmatrix} y^{(gks)} \\ 1 \end{bmatrix} \Rightarrow y^{(o)}$$

Eğitim kümesindeki *k.* veri için hesaplanan toplam ani hata

Çıkış katmanındaki j. nöron ile gizli katmandaki i. nörona ilişkin ağırlığın güncellenmesi



$$v_j^{(k)} = \sum_{i=0}^{m} w_{ji}^{(k)} y_i^{(k)} \qquad y_i^{(k)} = \varphi\left(v_j^{(k)}\right)$$

Hangi yöntem?

Ağırlığın güncellenmesi

$$w_{ji}^{(k+1)} = w_{ji}^{(k)} - \eta \frac{\partial \mathrm{E}^{(k)}}{\partial w_{ji}^{(k)}} \rightarrow$$

"en dik iniş "
(steepest descent)

$$E^{(k)} = \frac{1}{2} \sum_{j=1}^{q} e_j^{(k)^2} = \frac{1}{2} \sum_{j=1}^{q} (y_{d_j}^{(k)} - y_j^{(k)})^2 = \frac{1}{2} \sum_{j=1}^{q} (y_{d_j}^{(k)} - \varphi(\sum_{i=1}^{m} w_{ji} y_i^{(k)}))^2$$

Notasyona Dikkat!!!!! k eğitim kümesindeki kaçıncı veri olduğu aynı zamanda güncellemede bir iterasyon içinde kaçıncı defa güncellendiği

$y_j^{(k)}$ çıkış katmanında j. nöron çıkışı

$y_i^{(k)}$ gizli katmandaki i. nöron çıkışı

Yeni notasyon

Çıkış katmanı

Gizli katmanın sayısı

$$E = \frac{1}{2} \sum_{j=1}^{q} e_j^2 = \frac{1}{2} \sum_{j=1}^{q} (y_{d_j} - y_j^{(o)})^2 = \frac{1}{2} \sum_{j=1}^{q} (y_{d_j} - \varphi(\sum_{i=1}^{m} w_{ji}^{(o)} y_i^{(gks)}))^2$$

$$\frac{\partial E}{\partial w_{ji}^{(o)}} = \frac{\partial E}{\partial e_j} \frac{\partial e_j}{\partial y_j^{(o)}} \frac{\partial y_j^{(o)}}{\partial v_j^{(o)}} \frac{\partial v_j^{(o)}}{\partial w_{ji}^{(o)}} \qquad \frac{\partial E}{\partial w_{ji}^{(o)}} = e_j (-1) \varphi'(v_j^{(o)}) y_i^{(gks)}$$

Hatırlatma

Gizli katman ve çıkış katmanındaki her nöron iki iş yapıyor:
(i) nöron çıkışındaki işareti nöron girişindeki işaretler cinsinden hesaplıyor,
(ii) gradyen vektörünü geriye yayılım için yaklaşık olarak hesaplıyor

$$\frac{\partial E}{\partial w_{ji}^{(o)}} = \underbrace{e_j (-1) \varphi'(v_j^{(o)})}_{-\delta_j^{(o)}} y_i^{(gks)}$$

Yerel gradyen

$$\delta_j^{(o)} \triangleq -\frac{\partial E}{\partial v_j^{(o)}} = e_j \varphi'(v_j^{(o)})$$

$$w_{ji}^{(o)}(k+1) = w_{ji}^{(o)}(k) - \eta \left.\frac{\partial E}{\partial w_{ji}^{(o)}}\right|_k = w_{ji}^{(o)}(k) + \eta \delta_j^{(o)}(k) y_i^{(gks)}(k)$$

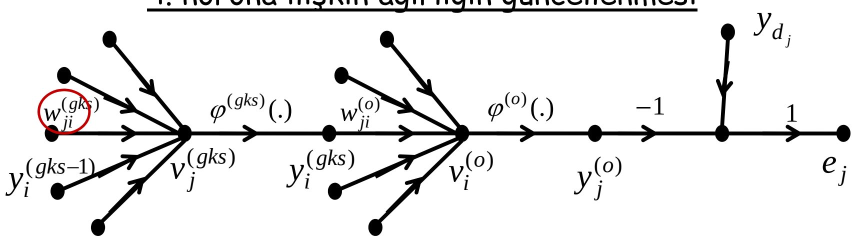
Çıkış katmanındaki tüm ağırlıkların güncellenmesi

$$w^{(o)}(k+1)=w^{(o)}(k)-\eta(-1)\left[\begin{matrix}\frac{\partial E}{\partial w_{11}^{(o)}} & \frac{\partial E}{\partial w_{12}^{(o)}} & \cdots & \frac{\partial E}{\partial w_{1m}^{(o)}}\\ \frac{\partial E}{\partial w_{21}^{(o)}} & \frac{\partial E}{\partial w_{22}^{(o)}} & \cdots & \frac{\partial E}{\partial w_{2m}^{(o)}}\\ \vdots & \vdots & & \vdots\\ \frac{\partial E}{\partial w_{q1}^{(o)}} & \frac{\partial E}{\partial w_{q2}^{(o)}} & \cdots & \frac{\partial E}{\partial w_{qm}^{(o)}}\end{matrix}\right]_k$$

$$w^{(o)}(k+1)=w^{(o)}(k)+\eta\begin{bmatrix}e_1\varphi'(v_1^{(o)}(k))\\ e_2\varphi'(v_2^{(o)}(k))\\ \vdots\\ e_q\varphi'(v_q^{(o)}(k))\end{bmatrix}\left[y_1^{(gks)}(k)\quad y_2^{(gks)}(k)\quad \ldots\quad y_m^{(gks)}(k)1\right]$$

$$w^{(o)}(k+1)=w^{(o)}(k)+\eta\begin{bmatrix}\delta_1^{(o)}(k)\\ \delta_2^{(o)}(k)\\ \\ \delta_q^{(o)}(k)\end{bmatrix}\left[y_1^{(gks)}(k)\quad y_2^{(gks)}(k)\quad \ldots\quad y_m^{(gks)}(k)1\right]$$

gizli katman (gks-1)'deki j. nöron ile gizli katman (o)'daki i. nörona ilişkin ağırlığın güncellenmesi



ilgilenilen $w_{ji}^{(gks)}$ ağırlığının toplam ani hataya etkisi

$$\mathrm{E}=\frac{1}{2}\sum_{j\in C}(e_j)^2=\frac{1}{2}\left[e_1\quad e_2\quad \ldots\quad e_q\right]\begin{bmatrix}e_1\\ e_2\\ \vdots\\ e_q\end{bmatrix}=\frac{1}{2}e^Te$$

Tanıdık birşeyler arayalım

$$\frac{\partial E}{\partial w_{ji}^{(gks)}}=\underbrace{\frac{\partial E}{\partial e_1}\frac{\partial e_1}{\partial y_1^{(o)}}\frac{\partial y_1^{(o)}}{\partial v_1^{(o)}}}_{-\delta_1^{(o)}}\underbrace{\frac{\partial v_1^{(o)}}{\partial y_j^{(gks)}}}_{w_{1j}^{(o)}}\frac{\partial y_j^{(gks)}}{\partial v_i^{(gks)}}\frac{\partial v_i^{(gks)}}{\partial w_{ji}^{(gks)}}+\underbrace{\frac{\partial E}{\partial e_2}\frac{\partial e_2}{\partial y_2^{(o)}}\frac{\partial y_2^{(o)}}{\partial v_2^{(o)}}}_{-\delta_2^{(o)}}\underbrace{\frac{\partial v_2^{(o)}}{\partial y_j^{(gks)}}}_{w_{2j}^{(o)}}\frac{\partial y_j^{(gks)}}{\partial v_i^{(gks)}}\frac{\partial v_i^{(gks)}}{\partial w_{ji}^{(gks)}}+\cdots+\underbrace{\frac{\partial E}{\partial e_q}\frac{\partial e_q}{\partial y_q^{(o)}}\frac{\partial y_q^{(o)}}{\partial v_q^{(o)}}}_{-\delta_q^{(o)}}\underbrace{\frac{\partial v_q^{(o)}}{\partial y_j^{(gks)}}}_{w_{qj}^{(o)}}\frac{\partial y_j^{(gks)}}{\partial v_i^{(gks)}}\frac{\partial v_i^{(gks)}}{\partial w_{ji}^{(gks)}}$$

$$\frac{\partial E}{\partial w_{ji}^{(gks)}}=-\delta_1^{(o)}w_{1j}^{(o)}\frac{\partial y_j^{(gks)}}{\partial v_i^{(gks)}}\frac{\partial v_i^{(gks)}}{\partial w_{ji}^{(gks)}}+-\delta_2^{(o)}w_{2j}^{(o)}\frac{\partial y_j^{(gks)}}{\partial v_i^{(gks)}}\frac{\partial v_i^{(gks)}}{\partial w_{ji}^{(gks)}}+\cdots-\delta_q^{(o)}w_{qj}^{(o)}\frac{\partial y_j^{(gks)}}{\partial v_i^{(gks)}}\frac{\partial v_i^{(gks)}}{\partial w_{ji}^{(gks)}}$$

$\varphi'(v_i^{(gks)})$ $\qquad$ $y_i^{(gks-1)}$

Yerel gradyen

$$\delta_j^{(gks)}\triangleq-\frac{\partial E}{\partial v_j^{(gks)}}=\sum_{i=1}^{q}w_{ij}\delta_i^{(o)}\varphi'(v_j^{(gks)})$$

$$w_{ji}^{(gks)}(k+1)=w_{ji}^{(gks)}(k)-\eta\left.\frac{\partial E}{\partial w_{ji}^{(gks)}}\right|_k=w_{ji}^{(gks)}(k)+\eta\delta_j^{(gks)}(k)y_i^{(gks-1)}(k)$$

Gizli katmanındaki tüm ağırlıkların güncellenmesi

$$\begin{bmatrix} \delta_1^{(gks)}(k) \\ \delta_2^{(gks)}(k) \\ \\ \delta_m^{(gks)}(k) \end{bmatrix} = \begin{bmatrix} w_{11}^{(o)} & w_{21}^{(o)} & \cdots & w_{q(m+1)}^{(o)} \\ w_{12}^{(o)} & w_{22}^{(o)} & \cdots & w_{q(m+1)}^{(o)} \\ \\ w_{1(m+1)}^{(o)} & w_{2(m+1)}^{(o)} & & w_{q(m+1)}^{(o)} \end{bmatrix} \begin{bmatrix} \delta_1^{(o)}(k) \\ \delta_2^{(o)}(k) \\ \\ \delta_q^{(o)}(k) \end{bmatrix} \times \begin{bmatrix} \varphi'(v_1^{(gks)}) \\ \varphi'(v_2^{(gks)}) \\ \\ \varphi'(v_m^{(gks)}) \end{bmatrix}$$

$$\delta^{(gks)} = w^{(o)^T} \delta^{(o)} \times \varphi'(v^{(gks)}) \quad \delta^{(gks)} = w^{(gks-1)^T} \delta^{(gks-1)} \times \varphi'(v^{(gks)})$$

$$w^{(gks)}(k+1) = w^{(gks)}(k)$$

Herhangi bir gizli katmandaki yerel gradyen

$$+\eta \begin{bmatrix} \delta_1^{(gks)}(k) \\ \delta_2^{(gks)}(k) \\ \vdots \\ \delta_m^{(gks)}(k) \end{bmatrix} \begin{bmatrix} y_1^{(gks-1)}(k) & y_2^{(gks-1)}(k) & \cdots & y_{ns\{gks-1\}}^{(gks-1)}(k) 1 \end{bmatrix}$$