

Geriyeye Yayılım Algoritması (Back-Propagation Algorithm)

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

Hesaplananlar: Eğitim Kümesindeki l. çiftte 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?

1

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

EK BİLGİ

Bazı Eniyileme (Optimizasyon) Teknikleri

Eniyileme problemi

$$\min_{x \in \Omega} f(x) \quad f(x): R^n \rightarrow R$$

$$\text{Kısıtlar: } h(x) = 0 \quad g(x) \leq 0$$

$$h(x): R^n \rightarrow R^p \quad g(x): R^n \rightarrow R^m$$

Kısıtsız Eniyileme Problemi

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

2

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

1. Mertebeden gerek koşul $x^* \in \Omega$ $f(\cdot)$ 'in ekstremum noktası ise $\nabla f(x^*) = 0$

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

2. Mertebeden yeter koşul (i) $\nabla f(x^*) = 0$

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

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

Doğrultu Belirleme (Line Search) Algoritması

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

3

Amaç: $f(x^{(k+1)}) < f(x^{(k)})$

Beklenti: Algoritma fonksiyonu enazlayan x^* 'a yakınsayacak

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

• s doğrultusunu belirle Nasıl ?

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

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

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

Bu doğrultuların işe yaradığını nasıl gösterebiliriz?

4

Gauss-Newton Metodu

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

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

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

5

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

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

Bu ifade aslında nedir?

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

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

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

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

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

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

Sonuç: x^* 'a yakınsayacak

EK BİLGİNİN SONU

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

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

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

7

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

Eğitim kümesindeki k . veri için ileri yolda hesaplanana yazalım:

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

1. Gizli Katman Çıkışı

$$\begin{bmatrix} v_1^{(2)} \\ v_2^{(2)} \\ \vdots \\ v_{ns2}^{(2)} \end{bmatrix} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & \dots & w_{1(ns1+1)}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & \dots & w_{2(ns1+1)}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{ns2}^{(2)} & w_{ns22}^{(2)} & \dots & w_{ns2(ns1+1)}^{(2)} \end{bmatrix} \begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \vdots \\ y_{ns1}^{(1)} \\ 1 \end{bmatrix} \xrightarrow{\varphi(\cdot)} \begin{bmatrix} y_1^{(2)} \\ y_2^{(2)} \\ \vdots \\ y_{ns2}^{(2)} \end{bmatrix}$$

2. Gizli Katman Çıkışı

8

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

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

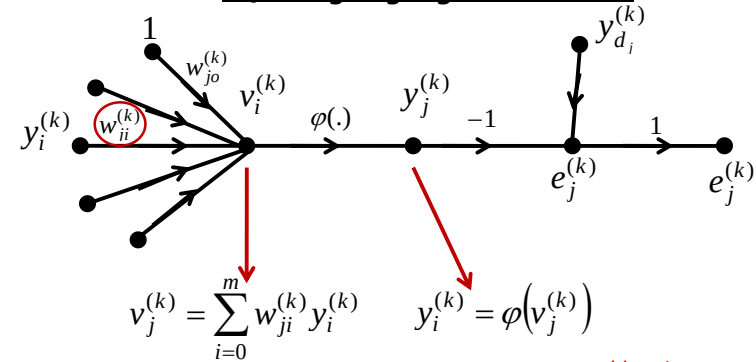
$$v^{(2)} = w^{(2)} \begin{bmatrix} y^{(1)} \\ 1 \end{bmatrix} \Rightarrow y^{(2)} \quad E^{(k)} = \frac{1}{2} \sum_{j \in C} (e_j^{(k)})^2 = \frac{1}{2} e^{(k)T} e^{(k)}$$

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

Eğitim kümesindeki k .
veri için hesaplanan
toplam ani hata

9

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



Hangi yöntem?

Ağırlığın güncellenmesi

$$w_{ji}^{(k+1)} = w_{ji}^{(k)} - \eta \frac{\partial E^{(k)}}{\partial w_{ji}^{(k)}} \rightarrow \text{"en dik iniş" (steepest descent)}$$

10

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

Notasyona Dikkat!!!! k eğitim kümesindeki kaçınıcı veri olduğu aynı zamanda güncellemede bir iterasyon içinde kaçınıcı 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

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

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

11

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

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

$-\delta_j^{(o)}$ Yerel gradyen

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

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

12

Çıkış katmanındaki tüm ağırlıkların güncellenmesi

$$w^{(o)}(k+1) = w^{(o)}(k) - \eta(-1) \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(o)}} & \frac{\partial E}{\partial w_{12}^{(o)}} & \dots & \frac{\partial E}{\partial w_{1m}^{(o)}} \\ \frac{\partial E}{\partial w_{21}^{(o)}} & \frac{\partial E}{\partial w_{22}^{(o)}} & \dots & \frac{\partial E}{\partial w_{2m}^{(o)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial E}{\partial w_{q1}^{(o)}} & \frac{\partial E}{\partial w_{q2}^{(o)}} & \dots & \frac{\partial E}{\partial w_{qm}^{(o)}} \end{bmatrix}_k$$

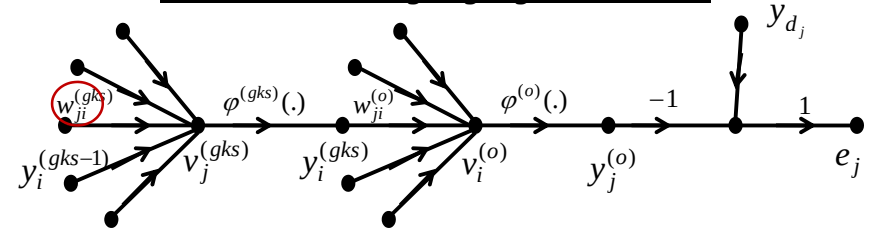
$$w^{(o)}(k+1) = w^{(o)}(k) + \eta \begin{bmatrix} e_1 \phi'(v_1^{(o)}(k)) \\ e_2 \phi'(v_2^{(o)}(k)) \\ \vdots \\ e_q \phi'(v_q^{(o)}(k)) \end{bmatrix} \begin{bmatrix} y_1^{(gks)}(k) & y_2^{(gks)}(k) & \dots & y_m^{(gks)}(k) 1 \end{bmatrix}$$

13

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

$$+ \eta \begin{bmatrix} \delta_1^{(o)}(k) \\ \delta_2^{(o)}(k) \\ \vdots \\ \delta_q^{(o)}(k) \end{bmatrix} \begin{bmatrix} y_1^{(gks)}(k) & y_2^{(gks)}(k) & \dots & y_m^{(gks)}(k) 1 \end{bmatrix}$$

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



14

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

$$E = \frac{1}{2} \sum_{j \in C} (e_j)^2 = \frac{1}{2} \begin{bmatrix} e_1 & e_2 & \dots & e_q \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_q \end{bmatrix} = \frac{1}{2} e^T e$$

Tanıdık birşeyler
arayalım

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

15

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

Yerel gradyen

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

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

16

Gizli katmanındaki tüm ağırlıkların güncellenmesi

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

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

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

Herhangi bir gizli
katmandaki yerel
gradyen

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