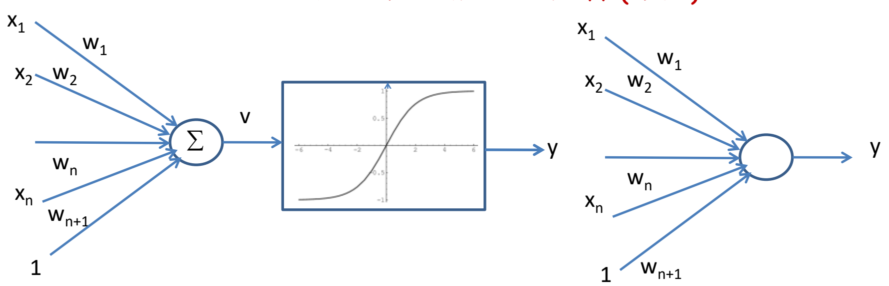
Genlikte Sürekli Algılayıcı (ADALİNE)

Bernard Widrow-Tedd Hoff (1960)



$$v = \begin{bmatrix} w_1 & w_2 & \dots & w_n & w_{n+1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \\ 1 \end{bmatrix}$$

$$v = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + w_{n+1} 1$$

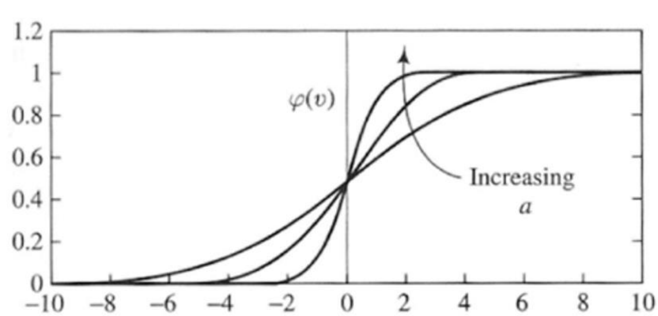
$$y = \varphi(v) = \tanh av$$

$$y = \varphi(v) = \frac{1}{1+e^{-av}}$$

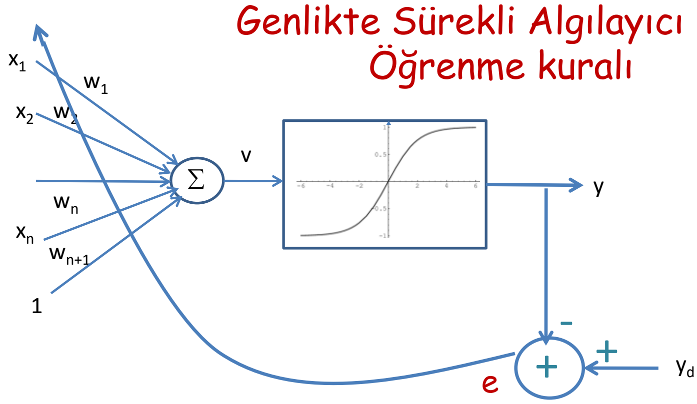
Genlikte Sürekli Algılayıcı (ADALİNE)

Bernard Widrow-Tedd Hoff (1960)

$$y = \varphi(v) = \frac{1}{1+e^{-av}}$$



Genlikte Sürekli Algılayıcı için Öğrenme kuralı



Amaç: hatayı azaltacak ağırlıkları belirlemek

Hataya ilişkin bir fonksiyon oluşturularak işe başlanacak

$$e = y_d - y$$

$$E = \frac{1}{2} e^T e = (y_d - y)^T (y_d - y) = (y_d - \varphi(v))^T (y_d - \varphi(v))$$

Nasıl azaltabiliriz?

Verilen eğitim kümesi için, hata fonksiyonu E 'yi 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: $h(x) = 0 \qquad \mathrm{g}(x) \leq 0$

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

Kısıtsız Eniyileme Problemi

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

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

1. Mertebeden gerek koşul

$x^* \in \Omega \quad f(.)$'in ekstremum noktası ise

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

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \qquad \nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

$$f(x_1, x_2, x_3) = x_1{}^2 + x_1 x_3 + x_2{}^2 x_3$$

$$\nabla f = \begin{bmatrix} 2x_1 + x_3 \\ 2x_2 x_3 \\ x_1 + x_2{}^2 \end{bmatrix} \qquad \nabla^2 f = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2x_3 & 2x_2 \\ 1 & 2x_2 & 0 \end{bmatrix} \text{ Hessian}$$

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

$$\nabla^2 f = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2x_3 & 2x_2 \\ 1 & 2x_2 & 0 \end{bmatrix} \qquad x^T H x > 0 \quad \forall x \in \mathbb{R}^n$$

$\rightarrow x^* \in \Omega \quad f(.)$'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

<u>Amaç:</u> $f(x^{(k+1)}) < f(x^{(k)})$

<u>Beklenti:</u> Algoritma fonksiyonu enazlayan x^*'a yakınsayacak

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

- 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

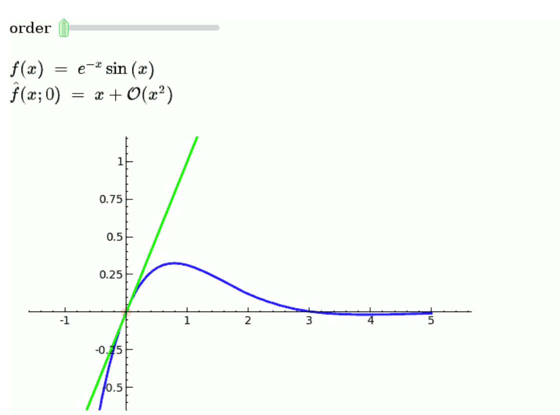
$* \alpha^{(k)} s^{(k)} \hat{=} -\frac{1}{2}[J^{(k)^T} J^{(k)}]^{-1} \nabla f(x^{(k)})$ → Gauss-Newton Metodu

Bu doğrultuların işe yaradığını nasıl gösterebiliriz?

Hatırlatma

f(x)'in x_0 civarında Taylor serisi açılımı.

$$f(x) = f(x_0) + \frac{1}{1!} f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \frac{1}{3!} f'''(x_0)(x - x_0)^3 + \cdots$$



"En dik iniş " (steepest descent) Metodu

$s^{(k)} \triangleq -\nabla f(x^{(k)})$ ile $f(x^{(k+1)}) < f(x^{(k)})$ sağlanır mı?

$f(x^{(k+1)})$ 'yı hesaplamanın bir yolu ne olabilir?

$f(x^{(k+1)})$ 'yı $x^{(k)}$ civarında Taylor serisine açalım.

$$f(x^{(k+1)}) \cong f(x^{(k)}) + \nabla f(x^{(k)})^T [x^{(k+1)} - x^{(k)}]$$

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} s^{(k)} \leftarrow s^{(k)} \triangleq -\nabla f(x^{(k)})$$

$$x^{(k+1)} - x^{(k)} = \alpha^{(k)} s^{(k)} \rightarrow x^{(k+1)} - x^{(k)} = -\alpha^{(k)} \nabla f(x^{(k)})$$

$$f(x^{(k+1)}) \cong f(x^{(k)}) - \alpha \nabla f(x^{(k)})^T \nabla f(x^{(k)})$$

$$f(x^{(k+1)}) \cong f(x^{(k)}) - \alpha \left\| \nabla f(x^{(k)}) \right\|^2$$

$$\rightarrow \quad f(x^{(k+1)}) < f(x^{(k)})$$

<u>Sonuç:</u> x^* 'a yakınsayacak

Özel durum: $f(x)$ Kuadratik $\rightarrow$ $f(x) = \frac{1}{2} x^T Q x - x^T b + c$

vektör

skaler

skaler

$$\nabla f(x) = Qx - b$$

1. Mertebeden gerek koşul $\quad \nabla f(x^*) = 0$

$$x^* = Q^{-1} b$$

$f(x)$ Kuadratik ise $\rightarrow x^{(k+1)} - x^{(k)} = 0$

Uygun $\alpha^{(k)}$ 'yı belirlemenin bir yolu var mı?

- $f(x^{(k)} + \alpha s^{(k)})$'yı α'ya göre enazlayan $\alpha^{(k)}$'yı belirle

$$\frac{d(f(x^{(k)} - \alpha^{(k)} \nabla f(x^{(k)}))}{d\alpha^{(k)}} = 0 \rightarrow \frac{d(f(x^{(k)} - \alpha^{(k)}(Qx^{(k)} - b))}{d\alpha^{(k)}} = 0$$

$$\rightarrow \quad \alpha^{(k)} = \frac{(Qx^{(k)} - b)^T (Qx^{(k)} - b)}{(Qx^{(k)} - b)^T Q (Qx^{(k)} - b)}$$

Newton Metodu

$\alpha^{(k)} s^{(k)} \triangleq -H^{-1}(x^{(k)}) \nabla f(x^{(k)})$ ile $f(x^{(k+1)}) < f(x^{(k)})$ sağlanır mı?

$f(x^{(k+1)})$'yı $x^{(k)}$ civarında Taylor serisine açalım.

$$f(x^{(k+1)}) \cong f(x^{(k)}) + \nabla f(x^{(k)})^T [x^{(k+1)} - x^{(k)}]$$
$$+ \frac{1}{2}[x^{(k+1)} - x^{(k)}]^T H(x^{(k)})[x^{(k+1)} - x^{(k)}]$$

$$x^{(k+1)} - x^{(k)} = \alpha^{(k)} s^{(k)} \leftarrow \alpha^{(k)} s^{(k)} \triangleq -H^{-1}(x^{(k)}) \nabla f(x^{(k)})$$

$$f(x^{(k+1)}) \cong f(x^{(k)}) + \nabla f(x^{(k)})^T [-H^{-1}(x^{(k)}) \nabla f(x^{(k)})]$$
$$+ \frac{1}{2}[-H^{-1}(x^{(k)}) \nabla f(x^{(k)})]^T H(x^{(k)})[-H^{-1}(x^{(k)}) \nabla f(x^{(k)})]$$

$$f(x^{(k+1)}) \cong f(x^{(k)}) - \nabla f(x^{(k)})^T H^{-1}(x^{(k)}) \nabla f(x^{(k)})$$
$$+ \frac{1}{2}[\nabla f(x^{(k)})^T H^{-1}(x^{(k)})^T \nabla f(x^{(k)})]$$

$$f(x^{(k+1)}) \cong f(x^{(k)}) - \frac{1}{2} \nabla f(x^{(k)})^T H^{-1}(x^{(k)}) \nabla f(x^{(k)})$$

$H(x^{(k)})$ Kesin Pozitif ise $\longrightarrow$ $f(x^{(k+1)}) < f(x^{(k)})$

Sonuç: x^*'a yakınsayacak

EK BİLGİNİN SONU

Amaç: $\min\limits_{w\in R^{n+1}} E$ Neden vazgeçtik?

$$E=\frac{1}{2}e^Te \;=(y_d-y)^T(y_d-y)=(y_d-\varphi(v))^T(y_d-\varphi(v))$$
$$=(y_d-\varphi(w_1x_1+w_2x_2+\ldots+w_nx_n+w_{n+1}1))^T$$
$$(y_d-\varphi(w_1x_1+w_2x_2+\ldots+w_nx_n+w_{n+1}1))$$

Böylece E'nin w'ya bağımlılığını açıkça yazdık, acaba $\min\limits_{w\in R^{n+1}} E$ sağlayan w'ları nasıl buluruz?

$E(w^{(k+1)})\le E(w^{(k)})$ sağlayacak $w^{(k+1)}$'i $w^{(k)}$ 'dan nasıl elde ederiz?

$$w^{(k)}+?=w^{(k+1)} \;\longrightarrow\; E(w^{(k+1)})\le E(w^{(k)})$$

$$w^{(k)}-c\nabla E=w^{(k+1)}$$

Nasıl bulunur?

$$E=\frac{1}{2}e^Te=\frac{1}{2}\,(y_d-y)^T(y_d-y)=\frac{1}{2}\,(y_d-\varphi(v))^T(y_d-\varphi(v))$$

$$v=w_1x_1+w_2x_2+\cdots+w_nx_n+w_{n+1}x_{n+1}$$

$$\nabla E \triangleq \begin{bmatrix}\frac{\partial E}{\partial w_1}\\ \frac{\partial E}{\partial w_2}\\ \bullet\\ \bullet\\ \bullet\\ \frac{\partial E}{\partial w_{n+1}}\end{bmatrix}=\underbrace{\frac{\partial E}{\partial e}}_{e}\overbrace{\frac{\partial e}{\partial y}}^{-1}\underbrace{\frac{\partial y}{\partial v}}_{\varphi'(v)}\frac{\partial v}{\partial w}$$

$$\frac{\partial v}{\partial w} \longrightarrow \begin{bmatrix}\frac{\partial v}{\partial w_1}\\ \frac{\partial v}{\partial w_2}\\ \bullet\\ \bullet\\ \bullet\\ \frac{\partial v}{\partial w_{n+1}}\end{bmatrix}=\begin{bmatrix}x_1\\ x_2\\ \bullet\\ \bullet\\ \bullet\\ 1\end{bmatrix}$$

$$\nabla E \triangleq \begin{bmatrix} \frac{\partial E}{\partial w_1} \\ \frac{\partial E}{\partial w_2} \\ \bullet \\ \bullet \\ \bullet \\ \frac{\partial E}{\partial w_{n+1}} \end{bmatrix} = \frac{\partial E}{\partial e}\frac{\partial e}{\partial y}\frac{\partial y}{\partial v}\frac{\partial v}{\partial w} = -(y_d - y)\varphi'(v)\begin{bmatrix} x_1 \\ x_2 \\ \bullet \\ \bullet \\ \bullet \\ 1 \end{bmatrix}$$

Öğrenme Kuralı:

$$\Delta w = -(y_d - y)\varphi'(v)x$$

Aktivasyon fonksiyonunu lineer alırsak ...

$$\Delta w = -(y_d - y)\varphi'(v)x \rightarrow \Delta w = -(y_d - y)x$$

Bir de girişleri normalize edersek ...

$$\Delta w = -(y_d - y)\varphi'(v)x \rightarrow \Delta w = -(y_d - y)\frac{x}{\|x\|}$$

ADALINE'ı nerede kullanabiliriz

- sınıflandırma probleminde

Eğitim aşamasından sonra $y = \begin{cases} 1 & \sum_{i=1}^{n} w_{ij}x_i > 0 \\ -1 & \sum_{i=1}^{n} w_{ij}x_i \leq 0 \end{cases}$

- Yaklaşık eğri uydurma (Lineer regresyon)

Aktivasyon Fonksiyonun Seçimi

Eğitim değerler kümesi (-1,+1) aralığında ise Aktivasyon fonksiyonu f(v) = tanh(av) alınır. a parametresinin değer aralığı (0, 1)
Eğitim değerler kümesi (0,+1) aralığında ise Aktivasyon fonksiyonu f(v) = 1/(1+exp(-av)) alınır. a parametresinin değer aralığı (0, 1)
Eğitim değerler kümesi (-5,+10) aralığında ise Aktivasyon fonksiyonu f(v) = tanh(av) alınır. Burada; - Eğitim kümesi normalize edilebilir. Veya - Aktivasyon fonksiyonu f(v) = 10*tanh(av) olarak alınabilir.

Aktivasyon Fonksiyonların 1. türevleri

Hyperbolic Tangent Function

$g(z) = tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$

$$\begin{aligned}
\frac{d}{dz}\left(\frac{e^z - e^{-z}}{e^z + e^{-z}}\right) &= \frac{e^z + e^{-z}}{(e^z + e^{-z})^2} d\left(e^z - e^{-z}\right) - \frac{e^z - e^{-z}}{(e^z + e^{-z})^2} d\left(e^z + e^{-z}\right) \\
&= \frac{\left(e^z + e^{-z}\right)\left(e^z + e^{-z}\right)}{(e^z + e^{-z})^2} - \frac{\left(e^z - e^{-z}\right)\left(e^z - e^{-z}\right)}{(e^z + e^{-z})^2} \\
&= \frac{\left(e^z + e^{-z}\right)^2 - \left(e^z - e^{-z}\right)^2}{(e^z + e^{-z})^2} \\
&= 1 - \left(\frac{e^z - e^{-z}}{e^z + e^{-z}}\right)^2 \\
&= 1 - tanh(z)^2
\end{aligned}$$

Sigmoid Function

$$\sigma(\mathrm{x}) = \frac{1}{1 + e^{-\mathrm{x}}}$$

$$\frac{d(\sigma(\mathrm{x}))}{dx} = \frac{0 * (1 + e^{-\mathrm{x}}) - (1) * \left(e^{-\mathrm{x}} * (-1)\right)}{(1 + e^{-\mathrm{x}})^2}$$

$$\frac{d(\sigma(\mathrm{x}))}{dx} = \frac{(e^{-\mathrm{x}})}{(1 + e^{-\mathrm{x}})^2} = \frac{1\text{-}1 + (e^{-\mathrm{x}})}{(1 + e^{-\mathrm{x}})^2} = \frac{1 + e^{-\mathrm{x}}}{(1 + e^{-\mathrm{x}})^2} - \frac{1}{(1 + e^{-\mathrm{x}})^2}$$

$$\frac{d(\sigma(\mathrm{x}))}{dx} = \frac{1}{1 + e^{-\mathrm{x}}} * \left(1 - \frac{1}{1 + e^{-\mathrm{x}}}\right) = \sigma(\mathrm{x})\left(1 - \sigma(\mathrm{x})\right)$$

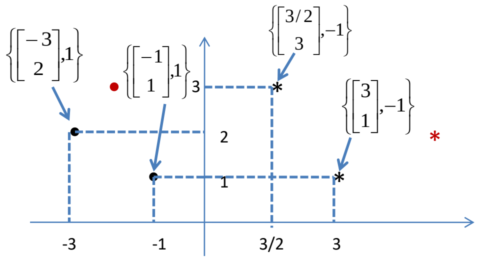
Örnek 1 (Sınıflandırma)

İlk ağırlık $w(1) = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$

Öğrenme hızı c=0.7

Durdurma ölçütü ε=0.01

$$\boldsymbol{tanh(x)} = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$$



x y_d

Eğitim kümesi: $\left(\left\{\begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}, 1\right\}, \left\{\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, -1\right\}, \left\{\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, 1\right\}, \left\{\begin{bmatrix} 3/2 \\ 3 \\ 1 \end{bmatrix}, -1\right\}\right)$

Test kümesi: $\left(\left\{\begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix}, -1\right\}, \left\{\begin{bmatrix} -3/2 \\ 3 \\ 1 \end{bmatrix}, 1\right\}\right)$

$$\left(\left\{\begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}, 1\right\}, \left\{\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, -1\right\}, \left\{\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, 1\right\}, \left\{\begin{bmatrix} 3/2 \\ 3 \\ 1 \end{bmatrix}, -1\right\}\right)$$

1. iterasyon

1. örüntü $v_1 = w^T(1)x_1 \Rightarrow [0 \quad -1 \quad 0]\begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix} = -2 \xrightarrow{\varphi} y_1 = -0.964$

$$w(2) = w(1) + c(y_{1d} - y_1)\varphi'(v_1)x_1$$

$$w(2) = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} + (0.7)\underbrace{(1 + 0.964)}_{\text{e}=\ 1.9640}(0.0707)\begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.2914 \\ -0.8057 \\ 0.0971 \end{bmatrix}$$

2. örüntü $v_2 = w^T(2)x_2 \Rightarrow w^T(2)\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} = -1.5828 \xrightarrow{\varphi} y_2 = -0.9190$

$$w(3) = w(2) + c(y_{2d} - y_2)\varphi'(v_2)x_2$$

$$w(3) = \begin{bmatrix} -0.2914 \\ -0.8057 \\ 0.0971 \end{bmatrix} + (0.7)\underbrace{(-1 + 0.9190)}_{\text{e}=-0.0810}(0.1554)\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.3178 \\ -0.8145 \\ 0.0883 \end{bmatrix}$$

$$\left(\left\{\begin{bmatrix}-3\\2\\1\end{bmatrix},1\right\},\left\{\begin{bmatrix}3\\1\\1\end{bmatrix},-1\right\},\left\{\begin{bmatrix}-1\\1\\1\end{bmatrix},1\right\},\left\{\begin{bmatrix}3/2\\3\\1\end{bmatrix},-1\right\}\right)$$

3. örüntü $v_3 = w^T(3)x_3 \Rightarrow w^T(3)\begin{bmatrix}-1\\1\\1\end{bmatrix} = -0.4084 \xrightarrow{\varphi} y_3 = -0.3871$

$$w(4) = w(3) + c(y_{3d} - y_3)\varphi'(v_3)x_3$$

$$w(4) = \begin{bmatrix}-0.3178\\-0.8145\\0.0883\end{bmatrix} + (0.7)\underbrace{(1+0.3871)}_{e=1.3871}(0.8501)\begin{bmatrix}-1\\1\\1\end{bmatrix} = \begin{bmatrix}-1.1433\\0.0109\\0.9138\end{bmatrix}$$

4. örüntü $v_4 = w^T(4)x_4 \Rightarrow w^T(4)\begin{bmatrix}3/2\\3\\1\end{bmatrix} = -0.7683 \xrightarrow{\varphi} y_4 = -0.646$

$$w(5) = w(4) + c(y_{4d} - y_4)\varphi'(v_4)x_4$$

$$w(5) = \begin{bmatrix}-1.1433\\0.0109\\0.9138\end{bmatrix} + (0.7)\underbrace{(-1+0.646)}_{e=-0.354}(0.5827)\begin{bmatrix}3/2\\3\\1\end{bmatrix} = \begin{bmatrix}-1.3599\\-0.4223\\0.7694\end{bmatrix}$$

1. İterasyon sonunda durdurma ölçütünün sağlayıp sağlamadığı kontrol edilir.

4 örüntüye ilişkin hataların $(y_d - y)$ kareleri alınır ve ortalaması hesaplanır.

$$e^T e = [1.9640 \quad -0.0810 \quad 1.3871 \quad -0.354]\begin{bmatrix}1.964\\-0.081\\1.3871\\-0.354\end{bmatrix} = 5.9134$$

Ortalama hata = 5.9134 / 4 = 1.4783

Hata_ortalama < $\varepsilon = 0.01$ koşulunu sağlamadığı için ikinci iterasyona geçilir.

2. iterasyon

1. örüntü $v_1 = w^T(5)x_1 \Rightarrow w^T(5)\begin{bmatrix}-3\\2\\1\end{bmatrix} = 4.0045 \xrightarrow{\varphi} y_1 = 0.9993$

$w(6) = w(5) + c(y_{1d} - y_1)\varphi'(v_1)x_1$

$$w(6) = \begin{bmatrix}-1.3599\\-0.4223\\0.7694\end{bmatrix} + (0.7)(1-0.9993)(0.0013)\begin{bmatrix}-3\\2\\1\end{bmatrix} = \begin{bmatrix}-1.3599\\-0.4223\\0.7694\end{bmatrix}$$

2. örüntü $w(7) = \begin{bmatrix}-1.3599\\-0.4223\\0.7694\end{bmatrix}$

3. örüntü $w(8) = \begin{bmatrix}-1.3654\\-0.4168\\0.7749\end{bmatrix}$

4. örüntü $w(9) = \begin{bmatrix}-1.3658\\-0.4175\\0.7747\end{bmatrix}$

2. İterasyon sonunda durdurma ölçütünün sağlayıp sağlamadığı kontrol edilir.

4 örüntüye ilişkin hataların $(y_d - y)$ kareleri alınır ve ortalaması hesaplanır.

$$e^Te = [0.0007 \quad -0.0011 \quad 0.0637 \quad -0.0128]\begin{bmatrix}0.0007\\-0.0011\\0.0637\\-0.0128\end{bmatrix} = 0.0042$$

Ortalama hata = 0.0042 / 4 = 0.00105

Hata_ortalama < $\varepsilon = 0.01$ koşulu sağladığından döngü sonlanır.

Ağın testi

w(9) ağırlık değerleri kullanılarak test küme örüntülerinin sonuçları kontrol edilir.

Test kümesi: $\left(\left\{\begin{bmatrix}5\\2\\1\end{bmatrix},-1\right\},\left\{\begin{bmatrix}-3/2\\3\\1\end{bmatrix},1\right\}\right)$

1. Örüntü için ağ sonucu = -0.999998 ✓
1. Örüntü için olması gereken = -1

2. Örüntü için ağ sonucu = 0.917154 ✓
2. Örüntü için olması gereken = 1

Hatırlatma

Eğitim aşamasından sonra $y = \begin{cases} 1 & \sum_{i=1}^{n} w_{ij} x_i > 0 \\ -1 & \sum_{i=1}^{n} w_{ij} x_i \leq 0 \end{cases}$

Örnek (Yaklaşık eğri uydurma - Lineer regresyon)

$$f(x_1, x_2) = 3x_1 + 2\cos x_2, \quad x_1 \in [0,1], \; x_2 \in [0, \pi/2]$$

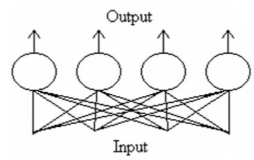
İlk ağırlık $w(0) = \begin{bmatrix}-0.1\\0.1\end{bmatrix}$ Öğrenme hızı c=0.7 ε=0.01

```
x1=rand(1000,1);
x2=floor(rand(1000,1)*(pi/2));
x=[x1 x2];
yd=3*x1+2*cos(x2);
```

Ağın eğitimini 5 örüntü için hesaplayınız.

2'den fazla sınıfa ayırmak istersek, nasıl ayıracağız?

Unutulmaması gereken kısıt ne?



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

Amaç: $y_d \in R^m$ iken bağıntıyı belirlemek

$$\begin{bmatrix} v_1^{(o)} \\ v_2^{(o)} \\ \vdots \\ v_m^{(o)} \end{bmatrix} = \begin{bmatrix} w_{11}^{(o)} & w_{12}^{(o)} & \cdots & w_{1(n+1)}^{(o)} \\ w_{21}^{(o)} & w_{22}^{(o)} & \cdots & w_{2(n+1)}^{(o)} \\ \vdots & \vdots & & \vdots \\ w_{m1}^{(o)} & w_{m2}^{(o)} & & w_{m(n+1)}^{(o)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ 1 \end{bmatrix}$$

Hataya ilişkin fonksiyon aynı ancak bu sefer tek hata yok, her bir çıkış için hata var: $e = y_d - y \in R^m$

$$E = \frac{1}{2} e^T e \; = (y_d - y)^T (y_d - y) = (y_d - \varphi(v))^T (y_d - \varphi(v))$$

$$= \left(\begin{bmatrix} y_{d1} \\ y_{d2} \\ \\ y_{dm} \end{bmatrix} - \varphi\left(\begin{bmatrix} w_{11}^{(o)} & w_{12}^{(o)} & \cdots & w_{1(n+1)}^{(o)} \\ w_{21}^{(o)} & w_{22}^{(o)} & \cdots & w_{2(n+1)}^{(o)} \\ \\ w_{m1}^{(o)} & w_{m2}^{(o)} & & w_{m(n+1)}^{(o)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_n \\ 1 \end{bmatrix} \right) \right)^T$$

$$\left(\begin{bmatrix} y_{d1} \\ y_{d2} \\ \\ y_{dm} \end{bmatrix} - \varphi\left(\begin{bmatrix} w_{11}^{(o)} & w_{12}^{(o)} & \cdots & w_{1(n+1)}^{(o)} \\ w_{21}^{(o)} & w_{22}^{(o)} & \cdots & w_{2(n+1)}^{(o)} \\ \\ w_{m1}^{(o)} & w_{m2}^{(o)} & & w_{m(n+1)}^{(o)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_n \\ 1 \end{bmatrix} \right) \right)$$

Öğrenme Kuralı: $\Delta w = -((y_d - y) \times \varphi'(v)) x^T$ ADALINE

2'den fazla sınıfa ayırmak istersek, nasıl ayıracağız?

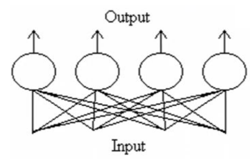
Amaç: m sınıfa ayırmak

$$S_1 : \sum_{i=1}^{n} w_i x_i + w_{n+1} < T_1$$

$$S_2 : T_1 \leq \sum_{i=1}^{n} w_i x_i + w_{n+1} < T_2$$

.....

$$S_m : T_{m-1} \leq \sum_{i=1}^{n} w_i x_i + w_{n+1} < T_m$$



Tek çıkışlıya göre farkı

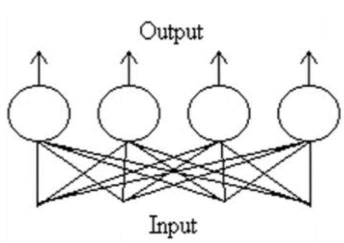
Öğrenme Kuralı: $$\Delta w_{m\times(n+1)} = c\frac{1}{2}[y_d - y]_{m\times 1}(x^T)_{1\times(n+1)}$$

Perceptron (GAA)

$$\Delta w_{m\times(n+1)} = -((y_d - y)_{m\times 1} \times \varphi'(v)_{m\times 1})(x^T)_{1\times(n+1)}$$

ADALINE

Örneğin 4 sınıfa ayrılacak ise 4 nöronlu bir ağ oluşturulur.



Sınıflar için hedeflenen vektörler şöyle oluşturulur

$$\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

Sınıf 1 Sınıf 2 Sınıf 3 Sınıf 4

√

Vektörlerin birbirine olan hamming mesafesi aynı

✗ $$\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}$$

Vektörlerin birbirine olan hamming mesafesi farklı

Hatırlatma

Her hangi 2 vektörün birbirine olan hamming mesafesi aynı

$\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix} \Rightarrow$ Hamming mesafesi 2

$\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix} \Rightarrow$ Hamming mesafesi 2

$\begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix} \Rightarrow$ Hamming mesafesi 2

Öneriler

Bölüm 3 (Okunması önerilen)

S. Haykin, "Neural Networks and Learning Machines", 3rd Edition, Pearson International Edition, 2009, New Jersey.

İzlenmesi önerilenler
----
https://www.youtube.com/watch?v=hc2Zj55j1zU

https://www.youtube.com/watch?v=skfNlwEbqck