

YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

57492

YAPAY SİNİR AĞLARININ DUYARLILIĞI

Matematik Müh. Nilgün Güler

F.B.E. Matematik Mühendisliği Anabilim Dalında

hazırlanan

YÜKSEK LİSANS TEZİ

Tez Danışmanı :Prof. Tahir Şişman

İSTANBUL, 1996

İÇİNDEKİLER

İÇİNDEKİLER.....	I
TEŞEKKÜR	II
TÜRKÇE ÖZET	III
ABSTRACT	IV
1.1.GİRİŞ	1
1.2.İleri Beslemeli Sinir Ağları.....	1
1.3.Geriye Yayılım Algoritması	3
2.1. AĞIRLIK HATALARINA KARŞI İLERİ BESLEMELİ SİNİR AĞLARININ DUYARLILIĞI	10
2.2. Hoff Hiperküre Alan Yakınsaması	16
2.3.Ağırlık Perturbasyonuna Göre Adaline Hatası	17
2.4. Girdi Hatasına Göre Karar Hatası	21
2.5. Ağırlık Perturbasyonu ve Giriş Hatasının Her İkisinde İçeren Karar hatası.....	23
2.6. Ağırlık Hatasından Kaynaklanan Madaline Çıkış Hatasının Olasılığının Hesaplanması	26
3. SİMULASYON SONUÇLARI	30
3.1 XOR Problemi	30
3.2. Gauss Sınıflama problemi	33
4. SONUÇ	35
KAYNAKLAR	36
ÖZGEÇMİŞ	

İÇİNDEKİLER

İÇİNDEKİLER.....	I
TEŞEKKÜR.....	II
TÜRKÇE ÖZET.....	III
ABSTRACT.....	IV
1.1.GİRİŞ.....	1
1.2.İleri Beslemeli Sinir Ağları.....	1
1.3.Geriye Yayılım Algoritması.....	3
2.1. AĞIRLIK HATALARINA KARŞI İLERİ BESLEMELİ SİNİR AĞLARININ DUYARLILIĞI.....	10
2.2. Hoff Hiperküre Alan Yakınsaması.....	16
2.3.Ağırlık Perturbasyonuna Göre Adaline Hatası.....	17
2.4. Girdi Hatasına Göre Karar Hatası.....	21
2.5. Ağırlık Perturbasyonu ve Giriş Hatasının Her İkisinde İçeren Karar hatası.....	23
2.6. Ağırlık Hatasından Kaynaklanan Madaline Çıkış Hatasının Olasılığının Hesaplanması.....	26
3. SİMÜLASYON SONUÇLARI.....	30
3.1 XOR Problemi.....	30
3.2. Gauss Sınıflama problemi.....	33
4. SONUÇ.....	35
KAYNAKLAR.....	36
EK1	
Bilgisayar programının çıktısı	
ÖZGEÇMİŞ	

TEŞEKKÜR

Yüksek lisans tezimin hazırlanması sırasındaki yardımlarından dolayı Yrd.Doç.Dr. Fikret Gürgen'e teşekkürlerimi bir borç bilirim.



Özet

Bu tezde; ileri beslemeli yapay sinir ağlarının, ağırlık hatası karşısındaki duyarlılıkları analiz edilmiştir. Ağ modeli olarak çok katmanlı algıyıcılar ve eğitim algoritması olarakta geriye yayılım algoritması kullanılmıştır

Çok katmanlı algılayıcılarda ağırlık vektörü giriş-çıkış dönüşümünü belirlemektedir. Dolayısıyla bu ağırlıklardaki her hangi bir değişiklik giriş-çıkış dönüşümümüzü etkilemektedir.

Yapay sinir ağıımızın duyarlılığını hesaplamak için artımsal ağırlık perturbasyon metodunu kullanıyoruz. Amacımız, artımsal bir ağırlık perturbasyonu karşısında karar hatası olasılığının hesaplanmasıdır. Bu amaçla iki tane yakınsama formülü geliştirdik: binormal yakınsama ve karekök yakınsaması.

Bunlardan binormal yakınsamanın doğruluk oranı çok yüksek olmasına rağmen; çok işlem gerektirdiğinden hesaplamalar için bilgisayara ihtiyaç duyuluyor ve ayrıca yakınsama, her katmandaki düğüm sayısına bağlıdır. Fakat katmandaki işlem elemanlarının sayısını arttırarak, hata probabilesi bir sayıya ulaştırılırsa yakınsamanın katmandaki işlem elemanından bağımsız olduğu görülür.

Karekök yakınsamasının sonucu binormal yakınsamasının sonucu kadar gerçek değere yakın değildir. Fakat çok daka kolay hesaplanabilir ve katmanda çok fazla işlem elemanı varsa; binormal yakınsamaya oldukça yakın sonuçlar verir.

ABSTRACT

The sensitivity of feedforward neural networks to weight errors is analyzed in this thesis. The network model used is the Multi Layer Perceptron (MLP) and the neural network is trained using Backpropagation algorithm.

The weight vector determines the input-output map of MLP. A slight change in the direction of the weight vector can alter this map .

Here, we use the method of additive weight perturbation to calculate sensitivity of our networks. Our goal is calculating probability of decision error. For this purpose, two approximations were derived which predict the probability of error for a single output of the network as a function of the percentage error in the weights: Binomial approximation and square root approximation.

The binomial approximation is the most accurate but requires evaluation by computer and depends on the network size.

The square root approximation is much easier to evaluate and is independent of the number of nodes. If there is enough nodes on the layer, the probability of error as predicted by the binomial approximation approaches the probability of error predicted by the square root approximation.

Bölüm :1

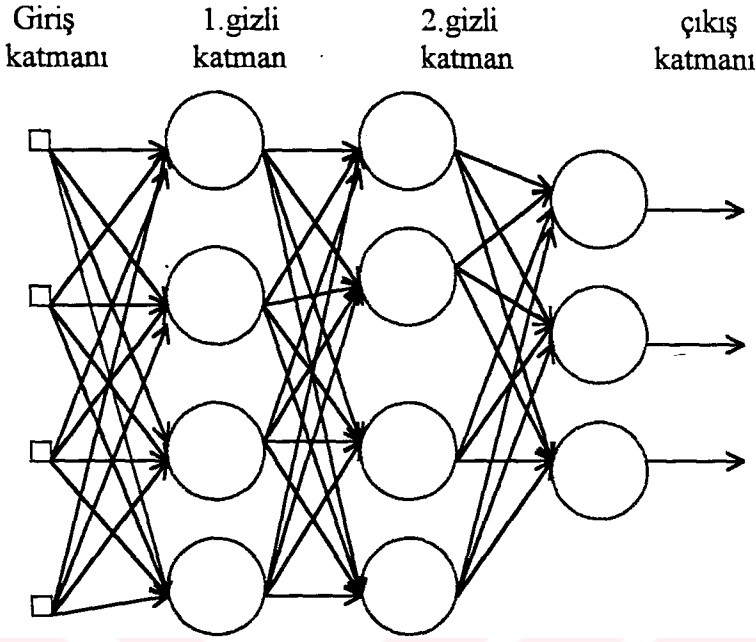
1.1. GİRİŞ

Yapay Sinir ağları (YSA), insan zekasının bazı fonksiyonlarını modellemek amacıyla, paralel olarak bir araya getirilmiş basit hesap birimleri dizileridir. Konuşma ve görüntü tanıma alanlarında, insan gibi davranış gösterebilmek için geniş bir işlem yeteneğine ihtiyaç duyulması nedeniyle, YSA'larına olan ilgi son yıllarda iyice artmıştır. Bu alanda yapılan çalışmalarla varılan sonuçlara göre; YSA'larında kullanılan ağ modeli, işlem elemanlarının bir araya bağlanma şekli ve ağırlıkların ayarlanması için kullanılacak öğrenme kuralının belirlenmesi çok önemlidir.

YSA'larında ağırlıklar; ağırlık giriş-çıkış dönüşümünü belirlemektedir. Dolayısıyla bu ağırlık değerlerinde yapılacak en ufak bir değişiklik ağırlık etkilemektedir. Bu tezde amacımız ağırlıklardaki değişiklikler karşısında ağırlık duyarlılığının hesaplanmasıdır.

1.2. İLERİ BESLEMELİ SİNİR AĞLARI

Bu bölümde YSA'larında çok önemli bir yer içeren ileri beslemeli çok katmalı ağların yapısına değinelim. Bu ağ, giriş değerlerinin tanıtıldığı giriş katmanı, bir yada daha fazla sayıda gizli katman ve çıkış işaretinin oluşturulduğu çıkış katmanından olmak üzere üç katmandan oluşmaktadır. Giriş işareti, ağ 'da bir katmandan bir sonraki katmana iletilecek şekilde ilerler. Bu tip YSA'a çok katmanlı algılayıcılar (ÇKA) denmektedir.



Şekil 1.1. Çok katmanlı algılayıcı

ÇKA, bir çok zor ve nonlinear problemlerin çözümlenmesinde başarıyla kullanılmaktadırlar. Eğitim algoritması olarak oldukça sık kullanılan ve hatanın düzeltilmesi kuralına dayanan, denetimli bir öğrenme çeşidi olan geriye yayılım algoritması kullanılır.

Geriye yayılım algoritması, ileri faz ve geri faz olmak üzere iki fazdan meydana gelmektedir. İleri fazda giriş vektörü ağ 'a tanıtılır ve etkisi ağ boyunca katman katman iletilirerek, çıkış katmanından çıkış değeri üretilir. Bu faz boyunca YSA'daki ağırlık değerleri değişmezdir. Geri faz'da ise hata düzeltme kuralı uygulanır. Bu kural, YSA'nın çıkış değeri ile istenilen çıkış değeri arasındaki fark hata değeri olarak adlandırılır, ve hata değeri YSA'da geriye doğru iletilerek ağırlık değerleri değiştirilir. Bu işlem istenilen hata değerine ulaşıncaya kadar tekrarlanır.

Çok katmanlı algılayıcılar üç belirgin özellik gösterir:

1) Yapay sinir ağındaki tüm işlem elemanları mutlaka nonlinear bir aktivasyon fonksiyonu içerirler. Bunu üzerinde önemle durulmasının nedeni nonlinear fonksiyonların her yerde türevi alınabilen fonksiyonlar olmasıdır. Ayrıca, eğer aktivasyon fonksiyonu nonlinear olamazsa, yapı kolaylıkla tek katmanlı algılayıcıya dönüştürülebilir.

2) Ağ, bir yada daha fazla sayıda gizli katman içermelidir ve bu katmanlar kesinlike giriş yada çıkış katmanlarının bir parçası olarak düşünülmemelidir. Bu gizli katmanlar aracılığıyla ağ çok kompleks yapıları bile kolaylıkla ile öğrenebilmektedir.

3) Ağ'da yüksek derecede bağlanabilirlik gerektirmektedir. Ağın bağlanabilirliğindeki bir değişiklik, ağırlık değerlerinin yada ağdaki bağlantı sayısının değiştirilmesini gerektirmektedir.

Bu özellikler ve öğrenme yeteneği biraraya gelince çok katmanlı algılayıcıların hesaplama gücü ortaya çıkmaktadır. Aynı özellikler, YSA'larının davranışları hakkındaki bilginizin eksik kalmasının nedenidir. Çünkü, nonlinnerliğin sınıflama formunun görüntüsü ve ağın yüksek bağlanabilirliği, çok katmanlı algılayıcıların teorik analizini güçleştirmektedir. Ayrıca, gizli katmanların kullanılması öğrenme işleminin izlenebilmesini zorlaştırmaktadır.

1.3. Geriye Yayılım Algoritması

YSA'daki giriş değerlerini x_j ($j=1,2,...,J$) ile çıkış değerlerini ise o_k ($k=1,2,...,K$) ile ifade edelim. J 'nci düğümün çıkışı ile k 'nci giriş'in arasındaki ağırlık değerlerini ise w_{kj} ile gösterelim.

Eğer vektör notasyonunu kullanacak olursak, ağıımızın ileri fazını aşağıdaki şekilde ifade edebiliriz

$$O = \Gamma [Wx] \quad (1.1)$$

Giriş, çıkış vektörleri ve ağırlık matrisleri

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \end{bmatrix}, \quad o = \begin{bmatrix} o_1 \\ o_2 \\ \vdots \\ o_k \end{bmatrix}$$

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1j} \\ w_{21} & w_{22} & \dots & w_{2j} \\ \vdots & \vdots & \vdots & \vdots \\ w_{k1} & w_{k2} & \dots & w_{kj} \end{bmatrix}$$

ve lineer olmayan $\Gamma [.]$ operatörü ise

$$\Gamma [.] = \begin{bmatrix} f(.) & 0 & \dots & 0 \\ 0 & f(.) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & f(.) \end{bmatrix}$$

ve hedef çıkış vektörü

$$d\Delta = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_k \end{bmatrix}$$

ile ifade edelim.

net_k k'nıncı katman için aktivasyon vektörüdür.

$$net_k = W \cdot x \quad (1.2)$$

Öklid hata ölçüsü

$$E_p = \frac{1}{2} \sum_{k=1}^K (d_{pk} - o_{pk})^2 = \frac{1}{2} \|d_p - o_p\|^2 \quad (1.3)$$

$$\Delta W_{kj} = -\eta \frac{\partial E}{\partial w_{kj}} \quad (1.4)$$

k katmanındaki her düğüm için $k=1,2,\dots,K$

$$net_k = \sum_{j=1}^J w_{kj} x_j$$

ve (1.1) formülünü kullanırsak, nöronların çıkışı

$$o_k = f(net_k) \text{ dır.} \quad (1.5)$$

Hata terimi olarak δ 'yı kullanacak olursak,

$$\delta_{ok} = -\frac{\partial E}{\partial (net_k)} \text{ dır.} \quad (1.6)$$

Zincir kuralını kullanarak $\partial E / \partial w_{kj}$ ifadesi aşağıdaki şekilde yazılabilir.

$$\frac{\partial E}{\partial w_{kj}} = \frac{\partial E}{\partial (net_k)} \frac{\partial (net_k)}{\partial w_{kj}} \quad (1.7)$$

(1.7)' in ikinci çarpan terimi ağırlıklar ve giriş örüntülerinin çarpımlarının toplamının türevidir.

$$\text{net}_k = w_{k1}x_1 + w_{k2}x_2 + \dots + w_{kj}x_j$$

$$\frac{\partial(\text{net}_k)}{\partial w_{kj}} = x_j$$

Bu ifadeyi denklem (1.7) de yerine yazarsak

$$\frac{\partial E}{\partial w_{kj}} = -\delta_{ok} x_j \text{ olur.}$$

İfadeyi (1.4) deki yerine koyarsak

$$\Delta W_{kj} = \eta \delta_{ok} x_j \quad k=1,2,\dots,K \quad j=1,2,\dots,J \quad (1.8)$$

Denklem (1.8) tek katmanlı YSA'larında delta eğitim/öğrenim ağırlık düzeltmesi formülüdür. Dikkat edilecek olursa formül herhangi bir aktivasyon formülü içermemektedir. δ_{ok} 'yı aktivasyon fonsiyonu içerecek şekilde yazabilmek için yeniden zincir kuralını kullanalım.

$$E(\text{net}_k) = E[O_k(\text{net}_k)]$$

$$\delta_{ok} = -\frac{\partial E}{\partial O_k} \frac{\partial O_k}{\partial(\text{net}_k)} \text{ dersek} \quad (1.9)$$

(1.5) de $O_k = f(\text{net}_k)$ demiştik. Öyleyse $f'(\text{net}_k) \equiv \frac{\partial O_k}{\partial(\text{net}_k)}$ dır.

$$\frac{\partial E}{\partial O_k} = \frac{\partial}{\partial O_k} \left(\frac{1}{2} \sum_{k=1}^K (d_{pk} - O_{pk})^2 \right)$$

$$= - (d_{pk} - o_{pk})$$

$$\delta_{ok} = (d_{pk} - o_{pk}) f'(net_k) \quad (1.10)$$

$$\Delta W_{kj} = \eta (d_{pk} - o_{pk}) f'(net_k) x_j \text{ olur.}$$

$$W_{kj}^{yeni} = W_{kj}^{eski} + \Delta W_{kj} \text{ dir.}$$

$f'(net_k)$ 'nin nonlinear ve türevi alınabilen bir aktivasyon fonksiyonu olması gerekir. Sigmoid fonksiyonu göz önüne alınırsa

$$\begin{aligned} f(net_k) &= \frac{1}{1 + e^{-net_k}}, \quad f(net_k) = o_k \\ f'(net_k) &= \frac{e^{-net_k}}{[1 + e^{-net_k}]^2} = \frac{1}{1 + e^{-net_k}} \frac{1 + e^{-net_k} - 1}{1 + e^{-net_k}} \\ &= \frac{1}{1 + e^{-net_k}} \left[\frac{1 + e^{-net_k}}{1 + e^{-net_k}} - \frac{1}{1 + e^{-net_k}} \right] \\ &= \frac{1}{1 + e^{-net_k}} \left[1 - \frac{1}{1 + e^{-net_k}} \right] \\ f'(net_k) &= o_k(1 - o_k) \end{aligned}$$

(1.10)'da yerine yazarsak

$$\delta_{ok} = (d_k - o_k) o_k (1 - o_k) \text{ olur.}$$

$$W'_{kj} = W_{kj} + \eta (d_k - o_k) o_k (1 - o_k) y_j$$

Formülü vektör notasyonunda yazacak olursak

$$W' = W + \eta \delta_o y'$$

δ_o hata terimlerini içeren sütun vektördür.

$$\delta_o \triangleq \begin{bmatrix} \delta_{o1} \\ \delta_{o2} \\ \vdots \\ \delta_{oK} \end{bmatrix}$$

δ_{ok} k'ninci nöronda d_k ile o_k arasındaki hata ifadesidir.

YSA'larında

gizli katman bulunduğu zaman aşağıdaki formül kullanılır:

ΔV_{ji} :Gizli katmandaki ağırlık değişimi olsun.

z_i : Giriş

$$\Delta V_{ji} = -\eta \frac{\partial E}{\partial V_{ji}}$$

$$\frac{\partial E}{\partial V_{ji}} = \frac{\partial E}{\partial (net_j)} \frac{\partial (net_j)}{\partial V_{ji}}$$

$$net_j = \sum_{i=1}^I V_{ji} z_i \text{ 'den } \frac{\partial (net_j)}{\partial V_{ji}} = z_i$$

$$\Delta V_{ji} = \eta \delta_{yj} z_i$$

$$\delta_{yj} \triangleq \frac{-\partial E}{\partial (net_j)}$$

$$\delta_{yj} = \frac{-\partial E}{\partial y_j} \frac{\partial y_j}{\partial (net_j)}$$

$$\frac{\partial E}{\partial y_j} = \frac{\partial}{\partial y_j} \left\{ \left\{ \frac{1}{2} \sum_{k=1}^K (d_k - f(net_k(y_j))) \right\}^2 \right\}$$

$$\frac{\partial y_j}{\partial (net_j)} = f'(net_j)$$

$$\frac{\partial E}{\partial y_j} = - \sum_{k=1}^K (d_k - o_k) y_j' \frac{\partial (net_k)}{\partial y_j} \quad (1.11)$$

$net_k = \sum_{j=1}^J w_{kj} y_j$ 'den $\frac{\partial net_k}{\partial y_j} = w_{kj}$ dir bunu (1.11) de yerine yazarsak

$$\begin{aligned} \frac{\partial E}{\partial y_j} &= - \sum_{k=1}^K \underbrace{(d_k - o_k)}_{\delta_{ok}} f'(net_k) w_{kj} \\ \frac{\partial E}{\partial y_j} &= f'(net_j) \sum_{k=1}^K \delta_{ok} w_{kj} \end{aligned}$$

Gizli katmanlar için ağırlık düzeltmesi

$$\Delta V_{ji} = \eta f_j'(net_j) z_i \sum_{k=1}^K \delta_{ok} w_{kj}$$

Gizli katmandaki yeni ağırlık değerini belirleyen formül;

$$V'_{ji} = V_{ji} + \eta f_j'(net_j) z_i \sum_{k=1}^K \delta_{ok} w_{kj} \text{ dir.}$$

BÖLÜM 2

2.1 AĞIRLIK HATALARINA KARŞI İLERİ BESLEMELİ SİNİR AĞLARININ DUYARLILIĞI

n boyutlu geometriden yararlanarak YSA'ımızı kolaylıkla analiz edebiliriz (Sommerwille, 1929). Bunun için Hoff hiperküre alan yakınsamasından yararlanacağız. Analize başlamadan önce tezde kullanılacak notasyon ve terimleri tanımlamaya çalışalım.

- $(n+1)$ boyutlu uzayda orijinden geçen X giriş vektörümüz $(x_0, x_1, x_2, \dots, x_n)$ olarak tanımlanmıştır.

- n boyutlu uzayda c noktasından r yarı çapı kadar uzaklıktaki noktaların oluşturduğu kürenin yüzey alanı:

$$A_n(r) = K_n r^{n-1} \text{ dir} \quad (2.1)$$

$$K_n = \frac{2\pi^{n/2}}{\Gamma(n/2)} \quad (2.2)$$

İspat: n boyutlu uzayda bir kürenin denklemi şudur:

$$x_1^2 + x_2^2 + \dots + x_n^2 = R^2$$

$r = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ dir. Ve kürede $f(r)$ sürekli fonksiyonu verilmiş olsun. Alan formülümüz n katlı integral cinsinden

$$\iint \dots \int f(r) dx_1 dx_2 \dots dx_n \text{ olur.}$$

Bu integralin hesaplanması için değişken dönüşümü yapalım.

$$\phi(x_1, x_2, \dots, x_n) = st.$$

$$\xi = \phi(x_1, x_2, \dots, x_n)$$

$$x_1 = X_1(\xi, \eta_1, \eta_2, \dots, \eta_{n-1})$$

$$x_1 = \eta_1$$

$$x_2 = X_2(\xi, \eta_1, \eta_2, \dots, \eta_{n-1})$$

$$x_2 = \eta_2$$

$$:$$

$$:$$

$$:$$

$$x_{n-1} = \eta_{n-1}$$

$$x_{n-1} = X_{n-1}(\xi, \eta_1, \eta_2, \dots, \eta_{n-1})$$

$$\iint \dots \int f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = \iint \dots \int \begin{vmatrix} \frac{\partial x_1}{\partial \xi} & \frac{\partial x_1}{\partial \eta_1} & \dots & \frac{\partial x_1}{\partial \eta_{n-1}} \\ \frac{\partial x_2}{\partial \xi} & \frac{\partial x_2}{\partial \eta_1} & \dots & \frac{\partial x_2}{\partial \eta_{n-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial \xi} & \frac{\partial x_n}{\partial \eta_1} & \dots & \frac{\partial x_n}{\partial \eta_{n-1}} \end{vmatrix} d\xi d\eta_1 \dots d\eta_{n-1}$$

$$\frac{\partial x_1}{\partial \xi} = 0$$

$$\frac{\partial x_1}{\partial \eta_1} = 1$$

$$\frac{\partial x_1}{\partial \eta_2} = 0$$

$$\frac{\partial x_2}{\partial \xi} = 0$$

$$\frac{\partial x_2}{\partial \eta_1} = 0$$

$$\frac{\partial x_2}{\partial \eta_2} = 1$$

$$:$$

$$:$$

$$:$$

$$\frac{\partial x_{n-1}}{\partial \xi} = 0$$

$$\frac{\partial x_{n-1}}{\partial \eta_1} = 0$$

$$\frac{\partial x_{n-1}}{\partial \eta_2} = 0$$

ξ 'ye göre türev alırsak

$$1 = \frac{\partial \phi}{\partial x_1} \frac{\partial x_1}{\partial \xi} + \frac{\partial \phi}{\partial x_2} \frac{\partial x_2}{\partial \xi} + \dots + \frac{\partial \phi}{\partial x_n} \frac{\partial x_n}{\partial \xi}$$

η_1 'in türevi

$$0 = \frac{\partial \phi}{\partial x_1} \frac{\partial x_1}{\partial \eta_1} + \frac{\partial \phi}{\partial x_2} \frac{\partial x_2}{\partial \eta_1} + \dots + \frac{\partial \phi}{\partial x_n} \frac{\partial x_n}{\partial \eta_1}$$

.....

.....

.....

$$0 = \frac{\partial \phi}{\partial x_1} \frac{\partial x_1}{\partial \eta_{n-1}} + \frac{\partial \phi}{\partial x_2} \frac{\partial x_2}{\partial \eta_{n-1}} + \dots + \frac{\partial \phi}{\partial x_n} \frac{\partial x_n}{\partial \eta_{n-1}}$$

türevlerin sonuçlarını değerlendirirsek;

$$1 = \frac{\partial \phi}{\partial x_1} \underbrace{\frac{\partial x_1}{\partial \xi}}_0 + \frac{\partial \phi}{\partial x_2} \underbrace{\frac{\partial x_2}{\partial \xi}}_0 + \dots + \frac{\partial \phi}{\partial x_n} \underbrace{\frac{\partial x_n}{\partial \xi}}_{\phi_{x_n}}$$

$$1 = \phi_{x_n} \frac{\partial x_n}{\partial \xi}, \text{ den } \frac{\partial x_n}{\partial \xi} = \frac{1}{\phi_{x_n}} \text{ dir.}$$

$$\phi_{x_1} + \phi_{x_n} \frac{\partial x_n}{\partial \eta_1} = 0 \text{ den } \frac{\partial x_n}{\partial \eta_1} = \frac{-\phi_{x_1}}{\phi_{x_n}}$$

.....

$$\frac{\partial x_n}{\partial \eta_{n-1}} = \frac{-\phi_{x_{n-1}}}{\phi_{x_n}}, \text{ dir.}$$

Bu sonuçları integral denklemimizde yerlerine yazarsak;

$$\begin{vmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \frac{1}{\phi_{x_n}} & \frac{-\phi_{x_1}}{\phi_{x_n}} & \frac{-\phi_{x_1}}{\phi_{x_n}} & \dots & \frac{-\phi_{x_1}}{\phi_{x_n}} \end{vmatrix} = \frac{1}{\phi_{x_n}} \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{vmatrix} = \frac{1}{\phi_{x_n}}$$

$$(dr^2)=2rdr$$

denklemimiz şu hali alır:

$$\int_0^R f(r) dr \underbrace{\int \dots dr}_{\Omega_n(r)} = \int_0^R f(r) \Omega_n(r) dr$$

$$\Omega_n(r) = 2r \int \dots \int \frac{dx_1 \dots dx_{n-1}}{x_n} \text{ dersek.} \quad x_v = r\xi_v \text{ 'den} \quad dx_v = r d\xi_v \text{ 'dir.}$$

$$dx_1 \dots dx_{n-1} = r^{n-1} d\xi_1 d\xi_2 \dots d\xi_{n-1}$$

gidip bunu Ω' da yerine yazarsak;

$$\Omega_n(r) = 2r^{n-1} \int \dots \int \frac{d\xi_1 d\xi_2 \dots d\xi_{n-1}}{\xi_n} = r^{n-1} w_n$$

$$w_n(r) = 2 \iint \dots \frac{d\xi_1 d\xi_2 \dots d\xi_n}{\xi_n}$$

$$\iint \dots \int f(r) dx_1 dx_2 \dots dx_n = w_n \int_0^r f(r) r^{n-1} dr \text{ dir.} \quad (2.3)$$

n katı kolaylıkla alınabilen, ve sağdan ile soldan integrali tek olan bir fonksiyon seçelim.

$$f(r) = e^{-(x_1^2 + x_2^2 + \dots + x_n^2)} = e^{-r^2}$$

Bunu gidip integralimizde $f(r)$ 'nin yerine yazalım.

$$\iint \dots \int e^{-(x_1^2 + x_2^2 + \dots + x_n^2)} dx_1 dx_2 \dots dx_n = \int_{-\infty}^{\infty} e^{-x_1^2} dx_1 \int_{-\infty}^{\infty} e^{-x_2^2} dx_2 \dots \int_{-\infty}^{\infty} e^{-x_n^2} dx_n$$

Bu ifadeyi aşağıdaki şekilde yazabiliriz:

$$\left(\int_{-\infty}^{\infty} e^{-x^2} dx \right)^n = \sqrt{\pi}$$

(2.3)'denklemini bu duruma göre yeniden yazarsak;

$$\underbrace{\left(\int_{-\infty}^{\infty} e^{-x^2} dx \right)^n}_{\sqrt{\pi}} = w_n \int_0^{\infty} e^{-r^2} r^{n-1} dr$$

$$\left(\sqrt{\pi} \right)^n = w_n \underbrace{\int_0^{\infty} e^{-r^2} r^{n-1} dr}_{\frac{1}{2} \Gamma(n/2)}$$

$$w_n = \frac{2\pi^{n/2}}{\Gamma(n/2)} \text{ dir.}$$

Bunu formüldeki yerine yazarsak , n boyutlu uzayda küre yüzeyinin alanının

$$A_n(r) = \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1} \text{ olduğunu ispatlamış oluruz.}$$

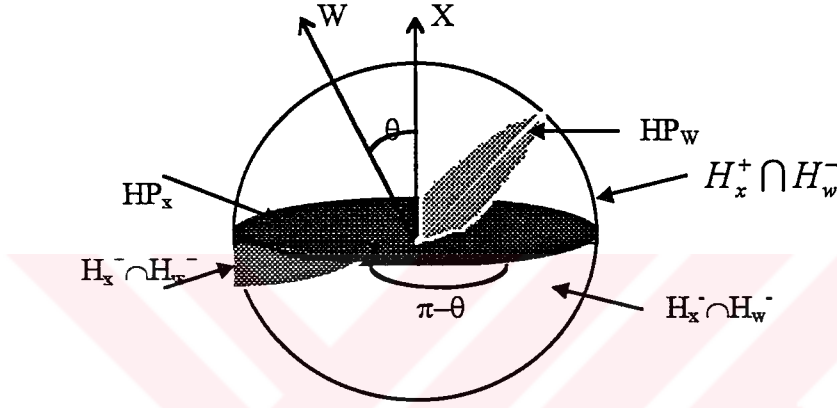
• X ve W ($W = [w_0, w_1, \dots, w_n]$) vektörlerinin skaler çarpımı c skaler sayısı olsun.
n boyutlu uzayda skaler sayı bir düzlemi gösterir.

$$X \cdot W \Delta \sum_{i=0}^n x_i w_i = c \quad (2.4)$$

Bu bir hiper düzlemdir. Ve bu düzlem W ağırlık vektörüne dik olup; düzlemin orijinden uzaklığı $c/|W|$ 'dir. HP_w (2.4)'de $c=0$ iken tanımlanan hiper düzlem olsun.

$$HP_w \triangleq \left\{ (x_0, x_1, \dots, x_n) : X \cdot W = \sum_{i=0}^n x_i w_i = 0 \right\} \quad (2.5)$$

• Hiper kürenin merkezinden geçen herhangi bir hiper düzlem küreyi homojen olarak iki eşit parçaya ayırır. Bu hiper düzlem HP_x ile gösterelim ve düzlem X vektörümüze dik olsun. Kürenin X 'in pozitif (+) olduğu yerlerini içeren kısmını H_x^+ ile X 'in negatif (-) olduğu yerleri içeren kısmını H_x^- notasyonu ile ifade edelim.



Şekil 2.1. Hiper küre.

• $(n+1)$ boyutlu uzayda X ve W vektörlerini gözönüne getirelim. HP_x hiper düzlemi küremizi H_x^+ ve H_x^- olarak homojen iki yarıya ayırırken; HP_w hiper düzlemide H_w^+ ve H_w^- olarak iki homojen küreye ayırır. (Şekil 2.1). θ , X vektörleri arasındaki açı olsun. θ 'yi $H_x^+ \cap H_w^+$ yada $H_x^- \cap H_w^-$ ile gösterebiliriz. Yada $(\pi-\theta)$ 'yi $H_x^+ \cap H_w^-$ veya $H_x^- \cap H_w^+$ ile ifade edebiliriz. θ açısının yüzeyinin, kürenin yüzeyine oranı $\theta/2\pi$ 'dir.

HP_w hiper düzleminin denklemi:

HP_w hiper düzleminin denklemi:

$$X \cdot W = \sum_{i=0}^n x_i w_i = 0 \text{ dır} \quad (2.6)$$

Bu hiper düzlem X 'i $X \cdot W > 0$ için ve $X \cdot W < 0$ için olmak üzere ikiye ayırır. Adaline'nın çıkış değeri giriş vektörü ile ağırlık vektörü arasındaki açıdan yararlanarak geometrik olarak hesaplanabilir; iki vektör arasındaki açı 90° 'den küçük ise çıkış değeri $+1$, 90° 'den büyük

ise çıkış değeri -1 değerini alır.

2.2. Hoff Hiperküre Alan Yakınsaması

Giriş değerleri ikili sayı sisteminde olduğundan sadece ± 1 değerlerini alacağından n giriş değişkenli Adaline'nımızın 2^n tane giriş örüntüsü mevcuttur. Bütün giriş örüntüleri n boyutlu uzayda yarı çapı $n^{1/2}$ olan merkezli kürenin üzerindeki noktalara karşılık gelmektedir. Bu yakınsamada n 'in oldukça büyük bir sayı olduğu ve giriş örüntülerimize karşılık gelen n boyutlu noktaların küre yüzeyinde düzgün bir dağılım gösterdiği kabul ediliyor. Bu yakınsamayı kısaca hiper kürenin seçilen bölgesindeki noktalara karşılık gelen giriş örüntülerinin yüzdesini; seçilen bölgenin yüzeyinin küre yüzeyinin tamamına oranına olarak özetleyebiliriz. Bu yakınmanın doğruluğu Hoff tarafından tezinde ispatlanmıştır (Hoff, 1960)

Bu hiper küre alan yakınsamasını Adaline'nımızın beklenen davranışını hesaplamakta kullanacağız. Bunun için yakınsamada küçük bir değişiklik yapmamız gerekmektedir. Bu değişiklik x_0 kutuplama değerinin her zaman $+1$ değerini almasından dolayı gerekmektedir. Bu da $(n+1)$ boyutlu giriş vektörünün her zaman kürenin üst yarısında dağılım göstereceği sonucunu çıkartmaktadır.

2.3. Ağırlık Perturbasyonuna Göre Adaline Hatası

Ağırlık vektörü, Adaline'nın giriş çıkış dönüşümünü belirlemektedir. Dolayısıyla bu vektördeki en ufak bir değişiklik bu dönüşümü değiştirmektedir. Burada, ağırlık vektöründeki perturbasyonun Adaline'nın giriş-çıkış dönüşümüne olan etkisini inceleyelim. Buradaki amacımız Adaline karar hatası olasılığının ağırlık perturbasyon oranı cinsinden ifade edilmesidir.

Bir Adaline elemanı ve bunun ağırlık (W) vektörü gözönüne alınsın. Eğer random olarak yönlendirilmiş ΔW perturbasyon vektörü W vektörüne eklenirse; elde edilen $W_p = W + \Delta W$ bizim pertube edilmiş ağırlık vektörüdür. Perturbasyon vektörünün büyüklüğünün, ağırlık vektörünün büyüklüğüne oranı bize ağırlık perturbasyon oranını (δW) verecektir.

$$\delta W = \frac{|\Delta W|}{|W|} \quad (2.7)$$

θ_{ww_p} , W ile W_p vektörleri arasındaki açısal perturbasyon olsun. İlk olarak, karar hatası olasılığının açısal perturbasyonla orantılı olduğunu gösterelim. Daha sonra da ağırlık perturbasyon oranı ile açısal ağırlık perturbasyonu arasındaki bağıntıyı kurmaya çalışırız.

$$P(\text{Karar Hatası}) = \frac{\theta_{ww_p}}{\pi} \quad (2.8)$$

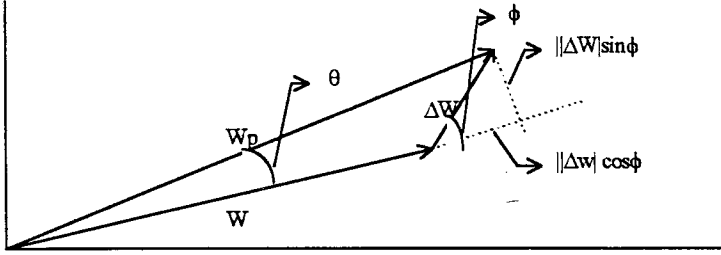
Karar hatası olasılığını, ağırlık perturbasyon oranı cinsinden ifade edebiliriz.

$$0 \leq \theta_{ww_p} \leq \sin^{-1}(\delta W)$$

θ_{ww_p} 'yi beklenen değer olan $\overline{\theta_{ww_p}}$ ile yer değiştirelim.

$$P(\text{Karar Hatası}) = \frac{\overline{\theta_{ww_p}}}{\pi} \quad (2.9)$$

Buradaki amacımız $\overline{\theta_{ww_p}}$ 'yi ağırlık perturbasyon oranı cinsinden bulmaktır. Bunun için Şekil 2.2'den yararlanılır.



Şekil 2.2. θ_{ww_p} geometrik gösterimi.

ϕ , W ile ΔW arasındaki bilinmeyen açıdır. Burada θ_{ww_p} 'yi $\phi, |W|, |\Delta W|$ cinsinden ifade edeceğiz.

$$\tan \theta_{ww_p} = \frac{|\Delta W \sin \phi|}{|W| + |\Delta W \cos \phi|} \quad (2.10)$$

$$\theta_{ww_p} = \tan^{-1} \left(\frac{|\Delta W \sin \phi|}{|W| + |\Delta W \cos \phi|} \right) \text{ ifadeyi } \frac{|W|}{|W|} \text{ ile çarparsak}$$

$$= \tan^{-1} \left(\frac{\frac{\overbrace{\Delta W}^{\delta W}}{|W|} \sin \phi}{\frac{|W| + |\Delta W \cos \phi|}{|W|}} \right)$$

$$= \tan^{-1} \left(\frac{\delta W \sin \phi}{1 + \delta W \sin \phi} \right) \quad \delta w \ll 1 \text{ için}$$

x 'in küçük değerleri için $\tan^{-1}x = x$ dir

$$\theta_{ww_p} = \delta w \sin \phi \quad (\delta w \text{'nin küçük değerleri için}) \quad (2.11)$$

Random θ_{w_p} değişkeni ağırlık perturbasyonu oranı ve random ϕ değişkeni aracılığıyla hesaplanabilir. Ağırlık perturbasyon oranının bilindiğini varsayarak θ_{w_p} 'nin beklenen değerini bulmaya çalışalım:

$$\overline{\theta_{w_p}} \approx \delta w E\{\sin \phi\} \quad \delta w \text{'nin küçük değerleri için} \quad (2.12)$$

(n+1) boyutlu uzayda random olarak yönlendirilmiş değişkenler arasındaki ϕ açısının olasılık yoğunluk fonksiyonu Glanz (1965) tarafından gösterilmiştir.

$$f_{\phi}(\phi) = (K_n/K_{n+1}) \sin^{n-1} \phi, \quad 0 < \phi < \pi \quad (2.13)$$

Burada $K_n = \frac{2\pi^{n/2}}{\Gamma(n/2)}$. Random ϕ değişkeninin olasılık yoğunluk fonksiyonunu kullanarak, n'nin büyük değerlerinde $\sin \phi$ 'nin beklenen değerinin bir'e oldukça yakın bir değer olduğu Winter (1989) tarafından bulunmuştur:

$$E\{\sin \phi\} = \int_0^{\pi} \sin \phi (K_n / K_{n+1}) \sin^{n-1} \phi d\phi = (K_n/K_{n+1})^2 2\pi/n \quad (2.14)$$

Bu fonksiyonu β (Beta) fonksiyonunu kullanarak çözmeye çalışalım:

$$\beta(x,y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt \quad (x \geq 1, y \geq 1) \text{ için}$$

Euler'in beta fonksiyonu $\sin \frac{\pi}{2}$ için $t=1$, $\sin 0$ için $t=0$ olur. $t=\sin^2 \phi$ dönüşümü yapalım. O

zaman $dt=2\sin \phi \cos \phi d\phi$ olur.

$$\begin{aligned}\beta(x,y) &= 2 \int_0^{\pi/2} \sin^{2x-2} \varphi \underbrace{(1 - \sin^2 \varphi)^{y-1}}_{\cos^2 \varphi} \sin \varphi \cos \varphi d\varphi \\ &= 2 \int_0^{\pi/2} \sin^{2x-1} \varphi \cos^{2y-1} \varphi d\varphi\end{aligned}$$

olacak $x = \frac{n+1}{2}$ ve $y = \frac{1}{2}$ değerlerini yerlerine yazalım. Bu durumda

$$\beta\left(\frac{n+1}{2}, \frac{1}{2}\right) = 2 \int_0^{\pi/2} \sin^n \varphi d\varphi \text{ olacak}$$

$\beta(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ ifadesinden yararlanarak

$$\begin{aligned}\beta\left(\frac{n+1}{2}, \frac{1}{2}\right) &= \frac{\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{n+1}{2} + \frac{1}{2}\right)} \\ &= \frac{\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma(1)}\end{aligned}$$

$$\int_0^{\pi/2} \sin^n \varphi d\varphi = \frac{1}{2} \beta\left(\frac{n+1}{2}, \frac{1}{2}\right) = \frac{1}{2} \frac{\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \text{ dir.}$$

$$E\{\sin \phi\} = \frac{K_n}{K_{n+1}} \int_0^{\pi} \sin^n \varphi d\varphi = 2 \left[\int_0^{\pi/2} \sin^n \varphi d\varphi \right] \frac{K_n}{K_{n+1}}$$

$$E\{\sin \phi\} = \frac{K_n}{K_{n+1}} \frac{\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \quad (2.15)$$

Başta $K_n = \frac{2\pi^{n/2}}{\Gamma(n/2)}$ olduğunu söylemiştik. Buna dayanarak

$$\frac{K_n}{K_{n+1}} = \frac{\frac{2\pi^{n/2}}{\Gamma(n/2)}}{\frac{2\pi^{(n+1)/2}}{\Gamma(\frac{n+1}{2})}} = \frac{1}{\pi} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}$$

$$\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} = \sqrt{\pi} \frac{K_n}{K_{n+1}} \quad (2.16)$$

(2.16) daki ifadeyi (2.15) deki yerine yazarsak

$$E\{\sin\phi\} = \left(\frac{K_n}{K_{n+1}}\right)^2 \pi \text{ bulunur.}$$

Stirling's yakınsamasını kullanırsak

$$\frac{K_n}{K_{n+1}} \approx \left(2\pi/n\right)^{-1/2}$$

n 'nin büyük değerleri için $E\{\sin\phi\} \approx 1$ dir ve δw 'nin küçük değerleri için $\overline{\theta_{ww_p}} \approx \delta w$ dir.

Bu sonucu gidip (2.9) eşitliğinde yerine koyarsak

$$P[\text{Karar hatası}] \approx \delta w / \pi. \quad (2.17)$$

Bu yakınsama n 'in oldukça büyük ve δw 'nin küçük değerleri için geçerlidir.

2.4 Girdi Hatasına Göre Karar Hatası

YNA'larında, bir katmandaki çıkış değeri bir sonraki katmanın giriş değeridir. Bu da bir katmandaki karar hatasının bir sonraki katmana giriş hatası olarak yansıtacağını gösterir. Bu yüzden input hatasının giriş-çıkış dönüşümüne olan etkisine inceleyelim.

x , $(n+1)$ boyutlu giriş vektörümüz olsun. x_p ise bu x vektörünü perturbe ederek elde ettiğimiz vektör olsun. Her iki vektöründe x_0 kutuplama bileşeni $+1$, diğer bileşenleride $+1$ veya -1 değerlerini almış olsunlar. Bu durumda iki vektör arasındaki fark *Hamming* uzaklığı cinsinden incelendiğinde, h *Hamming* uzaklığı $0 \leq h \leq n$ olabilir.

θ_{x_p} , x ve x_p vektörleri arasındaki açı olsun. θ_{x_p} 'nin π 'ye bölünmesi bize hata olasılığını vermektedir.

$$P(\text{Karar Hatası}) = \frac{\theta_{x_p}}{\pi} \quad (2.18)$$

θ_{x_p} 'yi iki vektörün skaler çarpım formülünü kullanarak kolayca bulabiliriz. Buna göre:

$$\begin{aligned} x \cdot x_p &= |x| |x_p| \cos \theta_{x_p} \\ \theta_{x_p} &= \cos^{-1} \left(\frac{x \cdot x_p}{|x| \cdot |x_p|} \right) \end{aligned} \quad (2.19)$$

x ve x_p vektörlerinin modülleri $(n+1)^{1/2}$ dir. Bu iki vektör'ün skaler çarpımları ise $n+1-2h$ dir. Buna göre;

$$\begin{aligned} \theta_{x_p} &= \cos^{-1} \left(\frac{n+1-2h}{n+1} \right) \\ \theta_{x_p} &= \cos^{-1} \left(1 - \frac{2h}{n+1} \right) \end{aligned} \quad (2.20)$$

$\cos \theta \approx 1 - \frac{\theta^2}{2}$ yakınsamasını kullanırsak

$$P(\text{Karar hatası}) \approx \frac{1}{\pi} \left[\frac{4h}{(n+1)} \right]^{1/2} \quad h \ll n+1 \quad (2.21)$$

$$X_p = X + \Delta X \quad (2.22)$$

x_i, x_{pi} ve Δx_i ; X, X_p ve ΔX vektörlerinin i 'inci bileşenleri bileşenleri olsun:

$$\Delta x_i = \begin{cases} 0, & x_{pi} = x_i \\ -2x_i & x_{pi} = -x_i \end{cases} \quad (2.23)$$

X ve X_p vektörleri arasındaki fark h kadardır. Öyleyse ΔX 'in büyüklüğü $(4h)^{1/2}$ dir. Böylelikle h giriş hatalı, δX girişperturbasyon oranı:

$$\delta X \triangleq \frac{|\Delta X|}{|X|} = \left[\frac{4h}{(n+1)} \right]^{1/2} \quad (2.24)$$

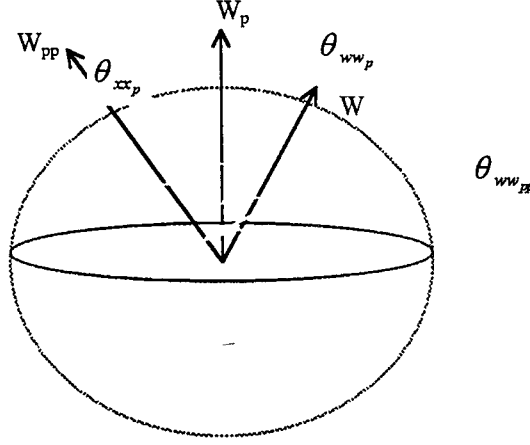
Bu ifadeyi h giriş hatalı karar hatası olasılığı için (2.21) deki yakınsamayla karşılaştırsak; Adaline karar hatası olasılığının δX giriş perturbasyon oranına bağlı olarak yakınsandığı görülür:

$$P(\text{Karar Hatası}) \approx \delta X / \pi \quad (2.25)$$

(2.17) ve (2.21) ifedelerini karşılaştırdığımızda karar hatası olasılığının perturbasyon oranının π 'ye bölünmesiyle ifade edildiğini görüyoruz. Bu da bize gösteriyorki giriş yada ağırlık perturbasyon oranlarının tanımlanmasında bir fark yoktur.

2.5. Ağırlık Perturbasyonu ve Giriş Hatasının Her İkisininide içeren Karar Hatası

Bundan önceki bölümde giriş hatası yada ağırlık perturbasyondan kaynaklanan hatanın probabilesini bulmuştuk. Eğer her iki vektörde (ağırlık ve giriş vektörleri) perturbe edersek, bu iki perturbasyonun toplam etkisini çıkaramayız. Bu yüzden hem ağırlık perturbasyonunu, hemde input hatasını içeren Adaline karar hatasının olasılığını bulan bir metod geliştirelim.



Şekil 2.3. Artımsal ağırlık perturbasyonu içeren giriş perturbasyonu.

Hem ağırlık perturbasyonunu hemde input perturbasyonunu içeren bir bir Adaline elemanı düşünelim. Bu iki perturbasyon miktarını pertürbe edilmiş ve pertürbe edilmemiş vektörler arasındaki açı cinsinden ifade edelim. θ_{xx_p} 'X ile X_p vektörleri arasındaki açı, θ_{ww_p} ise, W ile W_p vektörleri arasındaki açı olsun. Her iki perturbasyonun, hatanın olasılığına olan etkisi aynıdır. Giriş perturbasyonunu, ağırlık perturbasyonu üzerine yapılmış bir ekleme olarak düşünebiliriz. Buna göre , biz orijinal W ağırlık vektörünü θ_{xx_p} kadar döndürürsek elde edilen vektör W_p 'dir. (Şekil 2.3.) Bu vektörde θ_{xx_p} kadar döndürürsek iki kere pertürbe edilmiş olan W_{pp} vektörünü elde etmiş oluruz. Ağırlık vektörünün net açısal perturbasyonu $\theta_{ww_{pp}}$ ile verilmiştir. Toplam açısal perturbasyon $|\theta_{ww_p} - \theta_{xx_p}|$ ve $|\theta_{ww_p} + \theta_{xx_p}|$ 'dir. $\theta_{ww_{pp}}$ 'yi toplam açısal perturbasyonun beklenen değeri olduğunu düşünelim. Öyleyse

$$P(\text{Karar hatası}) = \frac{\overline{\theta_{ww_{pp}}}}{\pi} \quad (2.26)$$

Küçük perturbasyon oranları için açısal perturbasyon, perturbasyon oranına yakınsanabilir. Böylelikle toplam açısal perturbasyonunun beklenen değerini kullanarak toplam perturbasyon oranının beklenen değerini bulabiliriz. Bunun için random olarak

yönlendirilmiş P_1 ve P_2 vektörlerini düşünelim ve bunların büyüklükleri $(n+1)$ boyutlu uzayda değişmez olsun. ϕ , P_1 ve P_2 arasındaki random değerli açı, ve P' 'de P_1 ve P_2 'nin toplam vektörü olsun. P' 'nin büyüklüğü random değerli olsun.

$$|P| = (|P_1|^2 + |P_2|^2 - 2|P_1||P_2|\cos(\pi - \phi))^{1/2} \quad (2.27)$$

$|P|$ 'nin beklenen değeri; P_2 'nin büyüklüğünü ϕ 'nin olasılık yoğunluk fonksiyonuyla çarpımı yardımıyla bulunabilir.

$$E(|P|) = \int_0^\pi \underbrace{\left(\frac{K_n}{K_{n+1}} \right)}_{\propto \phi} \sin^{n-1} \phi (|P_1|^2 + |P_2|^2 - 2|P_1||P_2|\cos(\pi - \phi))^{1/2} d\phi \quad (2.28)$$

n 'nin büyük değerleri için $\sin^{n-1} \phi$ 0'a yakınsar, sadece ϕ 'nin $\pi/2$ 'ye yakın olduğu durumlar hariç. $\phi = \pi/2$ ise $\cos(\pi - \phi) = \cos(\pi/2) = 0$ 'dır. (2.28)'deki integral aşağıdaki hali alır:

$$E(|P|) \approx (|P_1|^2 + |P_2|^2)^{1/2} \int_0^\pi \underbrace{\left(\frac{K_n}{K_{n+1}} \right)}_1 \sin^{n-1} \phi d\phi \quad (2.29)$$

$$E(|P|) \approx (|P_1|^2 + |P_2|^2)^{1/2} \quad (2.30)$$

(2.30)'u kullanarak iki perturbasyondan kaynaklanan toplam perturbasyonun beklenen büyüklüğünü bulabiliriz. δx , giriş perturbasyon oranı; δw ağırlık perturbasyon oranı olsun.

$$E(\text{Toplam perturbasyon büyüklüğü}) \approx [(\delta X|W|)^2 + (\delta w|W|)^2]^{1/2} \quad (2.31)$$

$$E(\text{Toplam perturbasyon oranı}) \approx [(\delta X)^2 + (\delta W)^2]^{1/2} \quad (2.32)$$

Bu formülü yerine yazarsak ; δx giriş perturbasyon ve δw ağırlık perturbasyon değerlerine bağlı olarak Adaline karar hatası olasılığı:

$$P(\text{Karar hatası}) \approx \frac{1}{\pi} [(\delta X)^2 + (\delta W)^2]^{1/2} \text{ dir.} \quad (2.33)$$

2.6. Ağırlık Hatasından Kaynaklanan Madaniline Çıkış Hatasının Olasılığının Hesaplanması

Bu tezdeki amacımız, Adaline'nın ağırlık değerlerinde yapılan değişiklikler karşısındaki duyarlılığının bulunmasıdır. Bu amaçla referans ağ olarak, ağırlık değerleri random olarak seçilmiş bir Madaline ağı düşünelim. Bu ağı ağırlık vektörüne random değerli perturbasyon vektörünü ekleyerek pertürbe edilmiş ağıımızı oluşturalım. Perturbasyon vektörünün büyüklüğünün seçiminde dikkat edilecek nokta, ağıdaki bütün adaline elementlerinin değerlerinin aynı perturbasyon oranına sahip olmasıdır. Bu bölümde, seçilen pertürbe YSA'daki çıkış değerlerinin, orijinal YSA'daki çıkış değerlerine göre farklılıklarının olasılığını buluyoruz.

Buna göre Adaline yapısının birinci katmanı için hata olasılığını hesaplıyalım. Her iki ağıda aynı giriş değerleri tanıtılır. Birinci katmandaki hata kaynağı sadece ağırlık değerindeki perturbasyondur. Buna göre birinci katmandaki hata olasılığı yaklaşık olarak ağırlık perturbasyon oranının π 'ye bölümüdür.

$$PE_1 \approx \delta w / \pi \quad (2.34)$$

Birinci katmanın çıkış değeri, ikinci katmanın giriş değeridir. Bunun anlamı: ikinci ve sonraki katmanlardaki giriş değeri hataları, bir önceki katmanlardaki ağırlık perturbasyonundan kaynaklanmaktadır. $l > 1$ için, l katmanındaki Adaline elemanın hata olasılığı olan PE_l , ağırlık perturbasyon oranı ve o katmandaki giriş perturbasyon oranlarının her ikisinde fonksiyonudur. l katmanındaki giriş hata sayısı, $l-1$ katmanındaki çıkış hata sayısına eşittir. $l-1$ katmanındaki Adaline'nın ağırlık vektörünün bağımsız olduğu varsayılın, Adaline'nın çıkış vektöründe k tane hata olma olasılığı'nın hesaplanması binormal dağılıma uymaktadır. Buna göre:

$$P(l-1 \text{ katmanındaki } k \text{ tane çıkış hatası}) = \frac{n_{l-1}!}{k!(n_{l-1} - 1)!} (PE_{l-1})^k \cdot (1 - PE_{l-1})^{n_{l-1} - k} \quad (2.35)$$

PE_{lk} , perturbe edilmiş ağıımızın l katmanındaki Adaline'da meydana gelen k giriş hatasının şarta bağlı olasılığını gösterebilir. Buna göre k giriş hatasının giriş perturbasyon oranı $(4k/n_{l-1}+1)^{1/2}$ dir. (2.33) 'ü kullanarak

$$PE_{lk} = \frac{1}{\pi} \left[\frac{4k}{n_{l-1} + 1} + (\delta W)^2 \right]^{1/2} \quad (2.36)$$

$$PE_l = \sum_{k=0}^{n_{l-1}} PE_{lk} \cdot P(l-1 \text{ katmanındaki } k \text{ tane giriş hatası}) \quad (2.37)$$

Bu yakınsama (2.35) 'da binormal dağılım kullanıldığı için binormal yakınsama olarak adlandırılmış olsun. Dikkat edilecek olursa PE_L 'nin hesaplanabilmesi öncelikle $PE_1, PE_2, \dots, PE_{L-1}$ 'nin sırasıyla hesaplanması gerekmektedir. Bu da, hem PE_L 'nin analitik olarak ifade edilememesine, hemde bilgisayar kullanarak oldukça zahmetli hesaplamaların yapılmasını gerektirmektedir. Bunun için çok daha kolay hesaplanabilecek bir yakınsama çıkarmaya çalışalım.

Birinci katmanda n_1 tane Adaline elementi olsun. İlk katmandaki ağırlık perturbasyonu ikinci katmana giriş hatası olarak geleceğinden bu katmandaki giriş hatalarının beklenen sayısı $n_1 \cdot PE_1$ 'dir. Bunu formül (2.24) 'deki giriş hata sayısı olan h yerine $n_1 \cdot PE_1$ koyarsak aşağıdaki input yakınsamasını elde ederiz.

$$\delta X_{\text{katman2}} \approx \left[\frac{4PE_1 \cdot n_1}{n_1 + 1} \right]^{1/2} \quad (2.38)$$

n 'in oldukça büyük değerleri için $\left(\frac{n_1}{n_1 + 1} \right)^{1/2}$ yaklaşık 1'e gider öyleyse

$$\delta X_{\text{katman2}} \approx [4PE_1]^{1/2} \text{ yazabiliriz.} \quad (2.39)$$

Toplam perturbasyon oranı, giriş perturbasyon oranının karesiyle ağırlık perturbasyon oranının kareleri toplamının kareköküdür.

$$\left[2. \text{katmadaki toplam perturbasyon} \right] \approx \left[(\delta W)^{1/2} + 4PE_1 \right]^{1/2}$$

$$PE_2 \approx \frac{1}{\pi} [2. \text{Katmandaki toplam perturbasyon oranı}]$$

$$PE_2 \approx \frac{1}{\pi} \left[(\delta W)^2 + 4PE_1 \right]^{1/2}$$

$PE_1 \approx \delta w/\pi$ değerini yerine yazarsak:

$$\begin{aligned} PE_2 &\approx \frac{1}{\pi} \left[(\delta W)^2 + 4 \frac{\delta W}{\pi} \right]^{1/2} \\ &\approx \frac{1}{\pi} \delta W \left[1 + \frac{4}{\delta W \pi} \right]^{1/2} \end{aligned} \quad (2.40)$$

Bu yakınsama PE_3 'ün PE_2 'den bulunması, PE_4 'ün PE_3 'den bulunması şeklinde devam eder. Bu şekilde bir katmandan diğerine giden hata olasılığı bulunur. Buna göre 3'üncü katmandaki hata olasılığı

$$PE_3 \approx \frac{\delta W}{\pi} \left[1 + \frac{4}{\pi \delta W} \left(1 + \frac{4}{\pi \delta W} \right)^{1/2} \right]^{1/2} \quad (2.41)$$

4'üncü katmandaki hata olasılığı ise;

$$PE_4 \approx \frac{\delta W}{\pi} \left[1 + \frac{4}{\pi \delta W} \left(1 + \frac{4}{\pi \delta W} \left(1 + \frac{4}{\pi \delta W} \right)^{1/2} \right)^{1/2} \right]^{1/2} \quad (2.42)$$

$\frac{4}{\pi \delta W} = \beta$ dersek, 1 katmanındaki hata olasılığı:

PE_i 'deki β 'ların sayısı $l-1$ dir. Buna karek k yakınsaması denir. Karek k yakınsamasının hesaplanması kolay ve her katmandaki adaline eleman sayısından bağımsızdır.



BÖLÜM :3

3.1. SİMULASYON SONUÇLARI

Bir önceki bölümde, ileri beslemeli YSA'larının ağırlık değerlerindeki küçük perturbasyonlar karşısındaki duyarlıklarını hesaplamak için karekök yakınsaması ve binormal yakınsaması geliştirilmişti. Bu bölümde ise elde edilen teorik değerleri, bilgisayar simulasyon sonuçlarıyla karşılaştırarak, yakınsamaların doğruluklarının derecelerini görmeye çalışalım.

Bilgisayar progrmanında aktivasyon fonksiyonu olarak Hard-limited fonksiyonu kullanan ÇKA yerine, aktivasyon fonksiyonu sigmoid olan bir ÇKA yapısı kullanıldı. Bu iki fonksiyon arasındaki tezat., sigmoid fonksiyonu $\sum w_{ij}x_i=0$ iken çıkış değeri olarak 0.5 'i vermesidir. Sigmoid fonksiyonu, Hard-limited fonksiyonuna göre daha hassas bir sınıflama yapar.

3.1. XOR Problemi

Burada $\tanh(x)$ aktivasyon fonksiyonu nonlinear bir problem olan xor'a uygulanmıştır. YNA tek gizli katman içermektedir. Ağ'ın input giriş sayısı 2'dir.

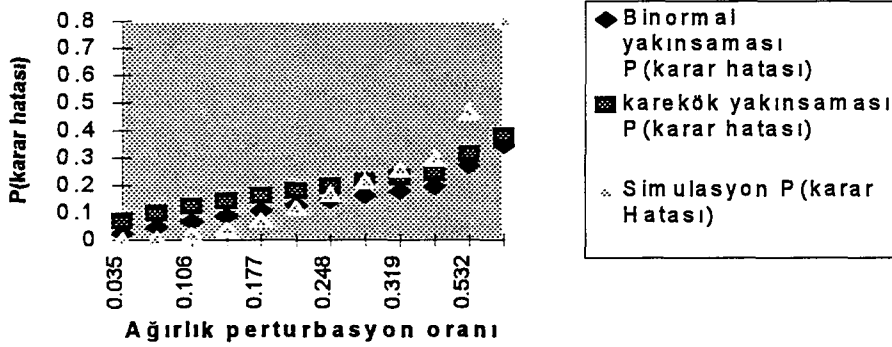
Xor problemini çözen ağımızın ağırlık perturbasyonuna olan duyarlılığını hesaplıyalım. İlk simulasyon programı nda ağdaki gizli katman 4 tane düğüme sahip olsun. itrasyon sayısı 500 olsun. Ağdaki ağırlık değerlerine (-2,+2) aralığında random artım miktarı eklensin. Bilgisayar programından elde edilen değerler aşağıda tablo 3.1'de verilmiştir.

Gizli katmandaki düğüm sayısı	4	iterasyon sayısı
Ağırlık perturbasyon miktarı	(-2.0-+2.0)	500

sw	Binormal yakınsaması P(karar hatası)	karekök yakınsaması P(karar hatası)	Simulasyon P(karar Hatası)
0.03547	0.023526842	0.068575805	0.000844158
0.07093	0.045892674	0.098286107	0.004864093
0.1064	0.067205131	0.121953078	0.0156583
0.14186	0.087564761	0.142617204	0.037223158
0.17732	0.107067284	0.161436698	0.071369533
0.21279	0.125800231	0.178993924	0.115792224
0.24825	0.143845434	0.195629035	0.165028715
0.28372	0.161278682	0.211559661	0.21336057
0.31918	0.178170033	0.226934215	0.25703205
0.35465	0.194584183	0.241858765	0.294765141
0.53197	0.27143877	0.3119314	0.465906361
0.7093	0.343741449	0.37746335	0.799031344

Table 3.1. (-2,+2) aralığındaki random perturbasyon sayısı için karar hatası olasılıkları

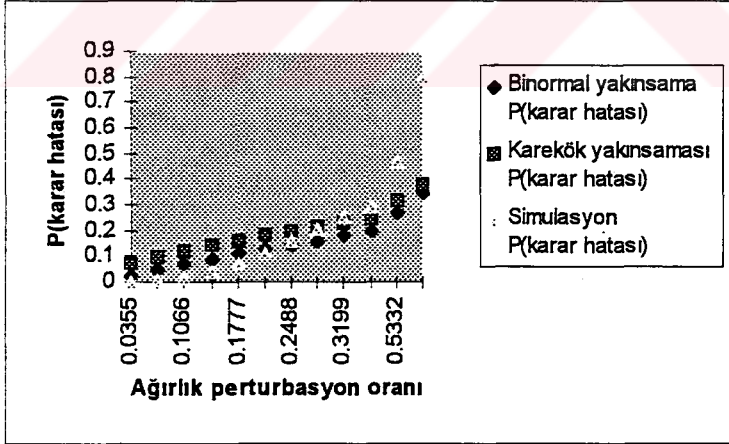
Tablo 3.1'e dayanarak aşağıdaki grafik çizilmiştir. Grafik'ten de görüldüğü gibi elde edilen yakınsamaların sonuçlarıyla bilgisayar programından elde edilen sonuçlar oldukça birbirine yakın.



3.1 Teorik yakınsamalar ile bilgisayar programından elde edilen sonuçların karşılaştırılması

Gizli katmandaki düğüm sayısı	6		iterasyon sayısı
Ağırlık perturbasyon miktarı	(-2.0-+2.0)		1000
Sw	Binormal yakınsama P(karar hatası)	Karekök yakınsaması P(karar hatası)	Simulasyon P(karar hatası)
0.035548	0.023580498	0.06865818	0.000749938
0.071097	0.045995418	0.098407855	0.004542919
0.106645	0.067351872	0.122107335	0.014830058
0.142194	0.08775285	0.142802399	0.035531659
0.177742	0.10729239	0.161650147	0.068583366
0.213291	0.12606027	0.179235203	0.111982055
0.248839	0.144137762	0.195897044	0.160538318
0.284388	0.161602118	0.21185444	0.208630167
0.319936	0.178522812	0.227255027	0.252450616
0.355485	0.194965845	0.242205827	0.290632205
0.533227	0.271960032	0.312407287	0.467996702
0.71097	0.344417292	0.378067663	0.801272246

Table 3.2. 4 düğümlü 10000 iterasyon



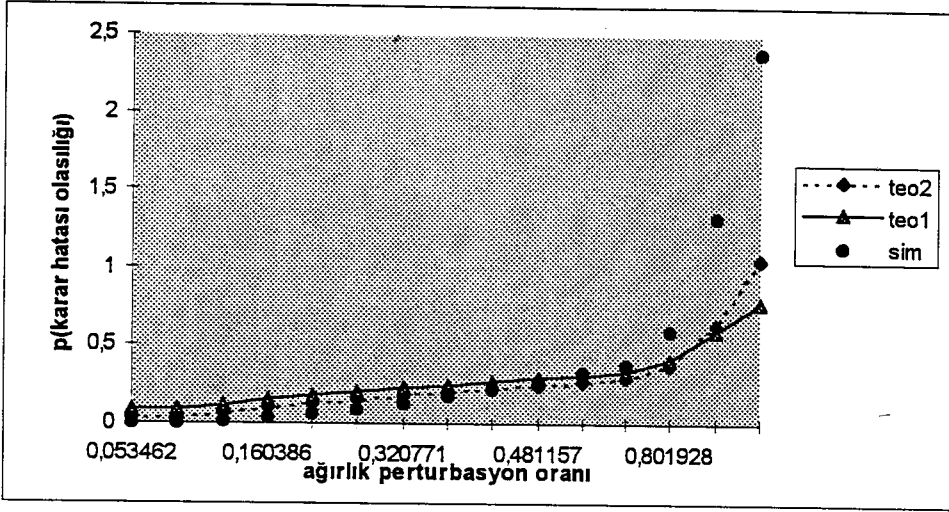
Şekil 3.2 4 düğümlü 1000 iterasyondan sonra

3.2. Gauss Sınıflama problemi

Gauss sınıflama probleminde 2 girişli ve iki çıkışlı ,bir tane gizli katman içeren bir ÇKA kullanılmıştır. Bu problemde ağ rastgele üretilen sayının hangi gauss eğrisine ait olduğunu buluyor. Her iki gauss dağılımı için varyans 1'e eşittir. Birinci dağılım için ortalama değeri 0 iken ,ikinci dağılım için 3'tür.

Gizli katmandaki düğüm sayısı	4
Ağırlık perturbasyon miktarı	(-4,+4)

Sw	Binormal yakınsama p(Karar hatası)	Karekök yakındaması P(karar hatası)	Simulasyon P(karar hatası)
0,053462	0,035014747	0,08477335	0,003857279
0,053462	0,035014747	0,08477335	0,003857279
0,106924	0,067515579	0,122279316	0,015686948
0,160386	0,097854194	0,152633977	0,035123268
0,213847	0,126348121	0,179502223	0,061586597
0,267309	0,153283626	0,204265307	0,094332726
0,320771	0,178914277	0,227611013	0,132511132
0,374233	0,203465639	0,249936832	0,175228956
0,427695	0,227137817	0,271494487	0,221608934
0,481157	0,250108881	0,292453897	0,270842243
0,534619	0,272538401	0,312935292	0,322232419
0,588081	0,29457094	0,333026879	0,375225922
0,801928	0,38127517	0,410624066	0,598031383
1,336547	0,629995693	0,594490253	1,326514115
1,871165	1,049262854	0,772103245	2,383648622



Grafikten de anlaşılaçağı üzere teorik sonuçlarla bilgisayar sonuçları oldukça yakın değerler vermektedir.

4. SONUÇ

Bu tezde ÇKA'ların giriş ve ağırlık değer hatalarına karşı duyarlıklarını hesaplamaya çalıştık. Ve önceki bölümde ağırlık hatası ile giriş hatasının kombinasyonundan oluşan karar hatası olasılığını hesaplamak için aşağıdaki yakınsamayı bulmuştuk.

$$P(\text{Karar hatası}) \approx \frac{1}{\pi} [(\delta X)^2 + (\delta W)^2]^{1/2}$$

Bu yakınsama ancak ağ oldukça büyük sayıda giriş değerine sahipse ve ağırlık perturbasyon oranı ile giriş perturbasyon oranları oldukça küçük değerler (0,5'den az) verilmişse, teorik ve deneysel sonuçlar hemen hemen aynı değerleri verir. Burada Adaline'nın duyarlılığı üzerine olan sonuçlar ÇKA'lara uygulanmıştır. Hesaplamalar sonunda binormal yakınsamanın gerçek değerlere oldukça yakın sonuçlar verdiği görülmüştür.

Binormal yakınsama gerçek değerlere oldukça yakın değerler veriyor olmasına rağmen, hesaplamalar için bilgisayar desteğine ihtiyaç duymaktadır. Ayrıca hesaplamalar ağın büyüklüğüne bağlıdır (katman sayısına ve her katmandaki düğüm sayısına).

Oysa karekök yakınsaması'nın hesaplanması çok daha kolaydır ve katmandaki Adaline sayısından bağımsızdır.

Bizim deney setlerimiz oldukça küçük olduğundan binormal yakınsamanın karekök yakınsamasına nazaran oldukça iyi sonuç verdiği görülür. Halbuki deneyler göstermektedirki; katmandaki ağırlık perturbasyon oranı ve katman sayısı sabit kaldığı sürece, katmandaki düğüm sayısı artarken binormal yakınsamanın probabilitate hatası, karekök yakınsamasının probabilitate hatasına erişir.

Kaynaklar

Glanz, F.H.; 1965. "Statistical extrapolation in certain adaptive pattern- recognition system ". Ph.D. dissertation, dept. Electrical Eng. Stanford Univ. Stanford.

Hoff, M.E., 1962. "Learning phenomena in networks of adaptive switching circuits", Ph.D. dissertation, dept. Electrical Eng. Stanford Univ. Stanford

Sommerville, D.M.Y., 1929. An Introduction to the Geometry of N Dimensions, London,England; Methuen and co.

Stevenson, M. and Winter, R. and Widrow B.,1990. "Sensitivity of Feedforward Neural Networks to Weight Errors". IEEE Transactions on Neural Networks. Vol. 1. No 1 pp 71-79

Widrow B. and Hoff, M.E, 1960. "Adaptive switching circuits" in IRE WESCON convention Record pt. 4. Pp.96-104

Widrow, B.,1962. "Generalization and information storage in networks of Adalines neurons in self organization system" pp. 435-460.

Widrow, B. and R.G. Winter,1988. "Neural nets for adaptive filtering and adaptive pattern recognition". IEEE computer Mag. pp 25-39

Winter, R.G.,1989. "Madaline Rule 2; A new method for traning networks of Adalines". Ph.D. dissertation, dept. Electrical Eng. Stanford Univ. Stanford CA.

ÖZGEÇMİŞ

Doğum Tarihi :24 Kasım 1971

Doğum Yeri :İstanbul

Eğitim: :1977-1982 Bahçelievler ilkokulu

1982-1985 Mühiddin Üstündağ İlköğretim Okulu

1985-1988 Ataköy Lisesi

1988-1992 Y.T.Ü. Matematik Mühendisliği

1992-1993 İ.T.Ü. Vakfi Hazırlık İngilizcesi

1993- .T.Ü. F.B.E. Yüksek lisans