

Geriye Yayılım Algoritması Bazı İpuçları

Öğrenme Hızı

- Öğrenme hızını belirleyen büyüklük η

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

- η küçük $\rightarrow$ ağırlıklardaki değişim bir iterasyondan diğerine küçük olacağı için, ağırlık uzayında düzgün bir değişim gözlenecek 😊
öğrenme yavaş olacak 😞
- η büyük $\rightarrow$ öğrenme hızlanacak 😊
salınım oluşacağından yakınsama mümkün olmayabilir 😞

Hızı artıralım ama salınım da olmasın. Bu mümkün mü?

Momentum Terimi

$$w_{ji}(k+1) = w_{ji}(k) + \eta \delta_j y_i + \underbrace{\alpha(w_{ji}(k) - w_{ji}(k-1))}_{\text{Momentum terimi}}$$

$$\Delta w_{ji}(k) \hat{=} w_{ji}(k+1) - w_{ji}(k) \rightarrow \Delta w_{ji}(k-1) = w_{ji}(k) - w_{ji}(k-1)$$

$$\Delta w_{ji}(k) = \alpha \Delta w_{ji}(k-1) + \eta \delta_j y_i \quad \hat{w}_{ji}(k) = \alpha \hat{w}_{ji}(k-1) + u(k)$$

Bu ifade neyi anımsatıyor?

$$x(k+1) = Ax(k) + Bu(k)$$

Lineer zamanla değişmeyen
ayrık zaman sistemi

HATIRLATMA

$$x(k+1) = Ax(k) + Bu(k)$$

$$x(1) = Ax(0) + Bu(0)$$

$$x(2) = Ax(1) + Bu(1)$$

$$x(2) = A[Ax(0) + Bu(0)] + Bu(1) \quad x(2) = A^2x(0) + ABu(0) + Bu(1)$$

$$x(2) = A^2x(0) + ABu(0) + Bu(1)$$

$$x(3) = A^3x(0) + A^2Bu(0) + ABu(1) + Bu(2)$$

$$\vdots$$

$$x(n) = A^n x(0) + \sum_{k=1}^n A^{(n-k)} Bu(k)$$

Bu sistemin çözümü nereye gidiyor?

- A matrisinin özdeğerleri birim daire içinde ise girişin belirlendiği deðere
- A matrisinin özdeğerleri birim daire üstünde ise salınım yapan bir sistem
- A matrisinin özdeğerleri birim daire dışında ise sonsuza

hatırlatmanın sonu

Momentum terimi varken güncellemede ne oluyor ona bakalım

$$\Delta w_{ji}(k) = \eta \sum_{t=0}^k \alpha^{n-t} \delta_j(t) y_i(t) \quad \Delta w_{ji}(k) = -\eta \sum_{t=0}^k \alpha^{n-t} \frac{\partial E(t)}{\partial w_{ji}(t)}$$

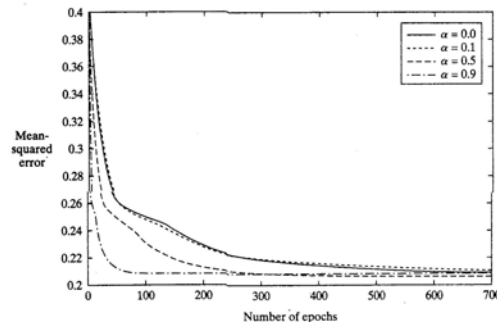
- $0 \leq |\alpha| < 1$

- Ardışık iterasyonlarda $\frac{\partial E(t)}{\partial w_{ji}(t)}$ aynı işaretli ise $\Delta w_{ji}(k)$ 'nin genliği büyüyecek, ağırlıklardaki değişim büyük olacak.

- Ardışık iterasyonlarda $\frac{\partial E(t)}{\partial w_{ji}(t)}$ farklı işaretli ise $\Delta w_{ji}(k)$ 'nin genliği azalacak, ağırlıklardaki değişim küçük olacak

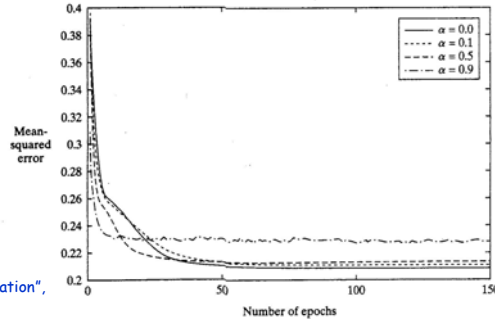


Momentum teriminin gradyenin işaretinin değiştiği doğrultularda kararlı kılma etkisi var.

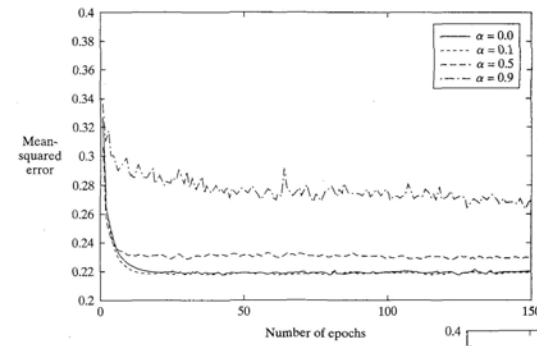


$\eta = 0.01$

$\eta = 0.1$

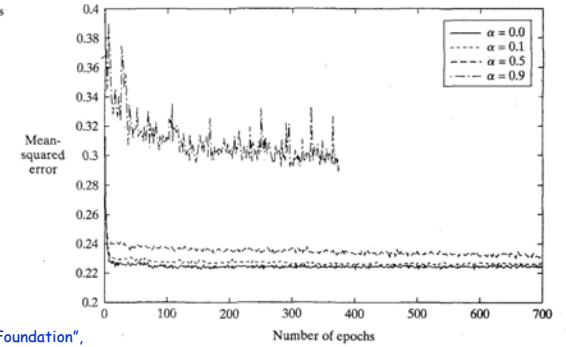


S. Haykin, "Neural Networks- A Comprehensive Foundation", 2nd Edition, Prentice Hall, 1999, New Jersey.

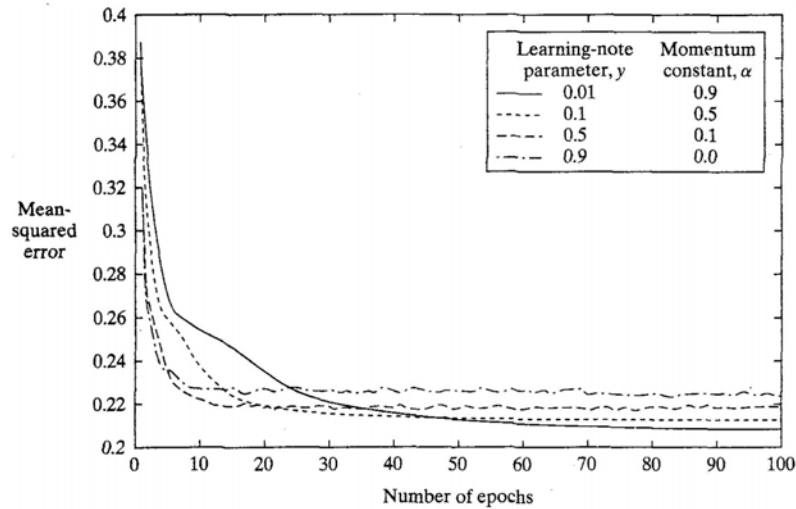


$\eta = 0.5$

$\eta = 0.9$



S. Haykin, "Neural Networks- A Comprehensive Foundation", 2nd Edition, Prentice Hall, 1999, New Jersey.



S. Haykin, "Neural Networks- A Comprehensive Foundation", 2nd Edition, Prentice Hall, 1999, New Jersey.

Adaptif Öğrenme Hızı $\eta_{ji}^{(k)}$

$$\Delta w_{ji}(k) = -\eta_{ji}^{(k)} \left[\frac{\partial E}{\partial w_{ji}} + \beta \Delta w_{ji}^{(k-1)} \right]$$

$$\eta_{ji}^{(k)} = \begin{cases} a \eta_{ji}^{(k-1)} & \frac{\partial E(w(k))}{\partial w_{ji}} \frac{\partial E(w(k-1))}{\partial w_{ji}} \geq 0 \\ b \eta_{ji}^{(k-1)} & \frac{\partial E(w(k))}{\partial w_{ji}} \frac{\partial E(w(k-1))}{\partial w_{ji}} < 0 \end{cases} \quad a \approx b^{-1}$$

$$\eta_{ji}^{(0)} = 10^{-3}, \quad \beta = 0.1, \quad 1.1 \leq a \leq 1.3, \quad 0.75 \leq b \leq 0.9$$

Grup-Veri Uyarlamalı Eğitim

- Veri Uyarlamalı Eğitim
"sequential mode"
"on-line mode"
"pattern mode"
"stochastic mode"

Eğitim kümesindeki her örüntü ağırlıklardan sonra ağırlıklar değiştiriliyor

$$w_{ji}(k+1) = w_{ji}(k) - \eta \frac{\partial E}{\partial w_{ji}} \Big|_k$$

- Grup Uyarlamalı Eğitim
"batch mode"

Eğitim kümesindeki tüm örüntüler ağırlıklardan sonra ağırlıklar değiştiriliyor

$$w_{ji}(k+1) = w_{ji}(k) - \eta \frac{\partial E_{ort}}{\partial w_{ji}}$$

	Grup Uyarlamalı	Veri uyarlamalı
Amaç Ölçütü	$w_{ji}(k+1) = w_{ji}(k) - \eta \frac{\partial E_{ort}}{\partial w_{ji}}$	$w_{ji}(k+1) = w_{ji}(k) - \eta \frac{\partial E}{\partial w_{ji}} \Big _k$
Her bağlantı için gereken bellek	☹️	😊
Örüntülerin ağırlıklara sunulması	☹️	😊
Algoritmanın yakınsaması	😊	☹️
Paralelliğin sağlanması	😊	☹️
Eğitim kümesinin fazlalıklı olması	☹️	😊
Algoritmanın basitliği	☹️	😊
Büyük boyutlu ve zor problemlerde etkin çözüm sağlanması	☹️	😊

Geriye Yayılım Algoritmasının Yakınsaması

Genlikte Ayırık Algılayıcıdaki gibi yakınsaması garanti değil.

Ne zaman durduracağız?

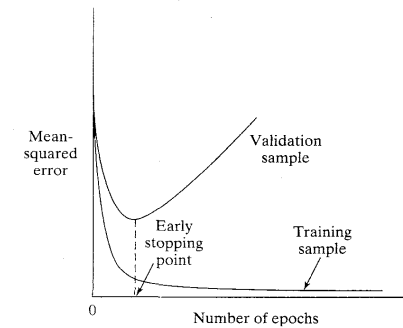
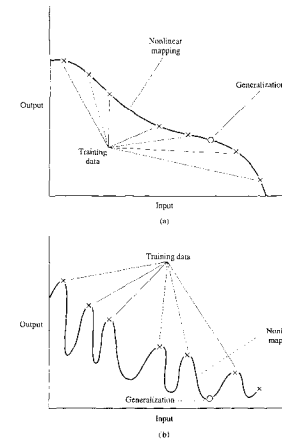
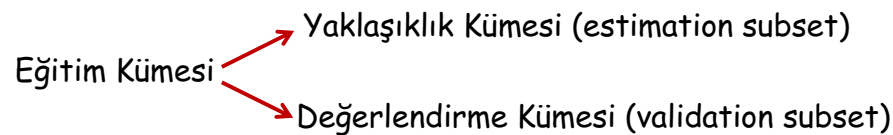
- Kramer+Sangionanni-Vincentelli (1989)

$$\|\nabla E\| \leq \varepsilon$$

$$|\Delta E_{ort}| \leq \varepsilon$$

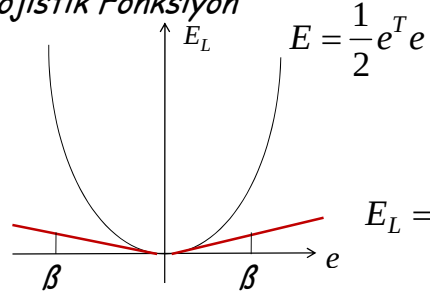
$$\max_j |e_j| \leq \varepsilon$$

- Çapraz değerlendirme (cross-validation)



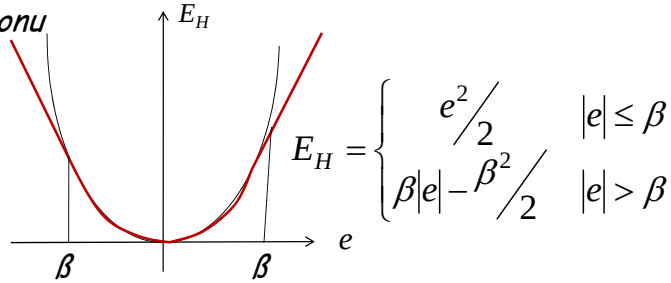
Hata Fonksiyonları

Lojistik Fonksiyon



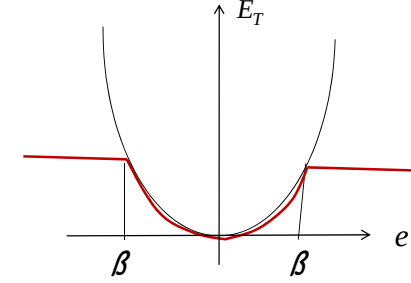
$$E_L = \beta^2 \ln[\cosh(e/\beta)]$$

Huber Fonksiyonu



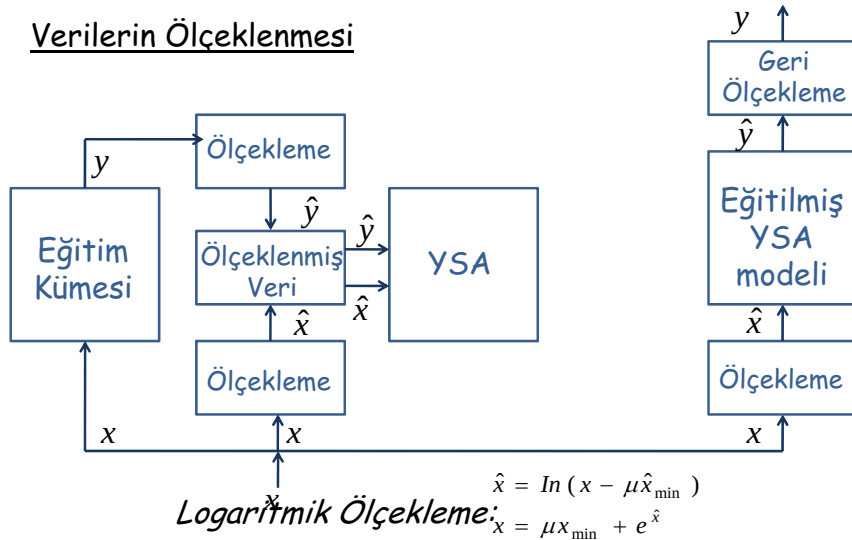
$$E_H = \begin{cases} e^2/2 & |e| \leq \beta \\ \beta|e| - \beta^2/2 & |e| > \beta \end{cases}$$

Talvar Fonksiyonu



$$E_T = \begin{cases} e^2/2 & |e| \leq \beta \\ \beta^2/2 & |e| > \beta \end{cases}$$

Verilerin Ölçeklenmesi



$$\hat{x} = \ln(x - \mu \hat{x}_{\min})$$

$$x = \mu x_{\min} + e^{\hat{x}}$$

Doğrusal Ölçekleme:

$$\hat{x} = \hat{x}_{\min} + \frac{x - x_{\min}}{x_{\max} - x_{\min}} (\hat{x}_{\max} - \hat{x}_{\min})$$

$$x = x_{\min} + \frac{\hat{x} - \hat{x}_{\min}}{\hat{x}_{\max} - \hat{x}_{\min}} (x_{\max} - x_{\min})$$

Son İpuçları

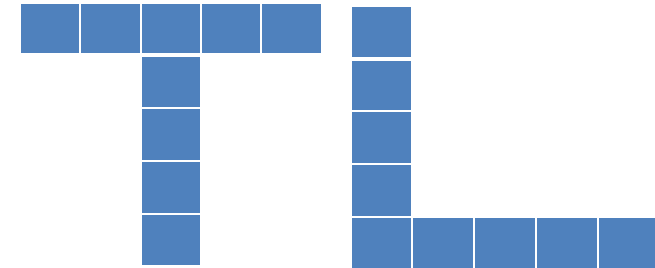
- $y_{d_i} \in \{\varepsilon_1, 1 - \varepsilon_2\}$, $x_i \in \{\varepsilon_1, 1 - \varepsilon_2\}$
 ε_i 'ler büyük ise ağırlıkların değişim aralığı küçülür; fiziksel gerçeklemeye uygun olur. Ancak girişlerdeki gürültüye tolerans azalır.
- Ağırlıklar başlangıçta aynı seçilirse, değişimleri de aynı olabilir. Dolayısıyla ağırlıklar yenilendiğinde aynı kalabilirler. Bunu engellemek için başlangıç ağırlıkları sıfıra yakın sayılardan rastgele seçilmeli.
- Ağırlıklar başlangıçta aynı seçilirse, değişimleri de aynı olabilir. Dolayısıyla ağırlıklar yenilendiğinde aynı kalabilirler. Bunu engellemek için başlangıç ağırlıkları sıfıra yakın sayılardan rastgele seçilmeli.
- Her katman eklendiğinde geriye yayılım algoritması yavaşlar. Bunu engellemek için girişten çıkışa doğrudan bağlantılar yapılabilir.

- Geriye yayılım algoritması "en dik iniş" ve gradyen yöneme dayalı olduğundan, bu yöntemi iyileştirici tüm teknikler geriye yayılım algoritmasını da iyileştirmek için kullanılabilir.

- İkinci türevleri kullanan lineer olmayan eniyileme yöntemlerinden herhangi biri de kullanılabilir.

SONLU ADIMDA GLOBAL MİNİMUMA YAKINSAMASI GARANTİ DEĞİLDİR.

T ve L harfini ayırt eden bir ağ

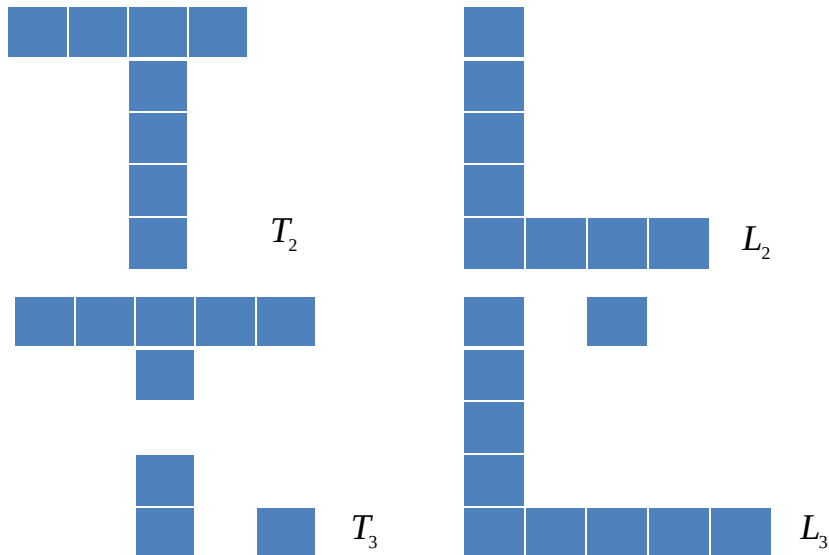


Bu harfleri ağa nasıl sunacağız? 25X1 boyutlu vektörler ile

$$T_1 = [1 \ 1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0]$$

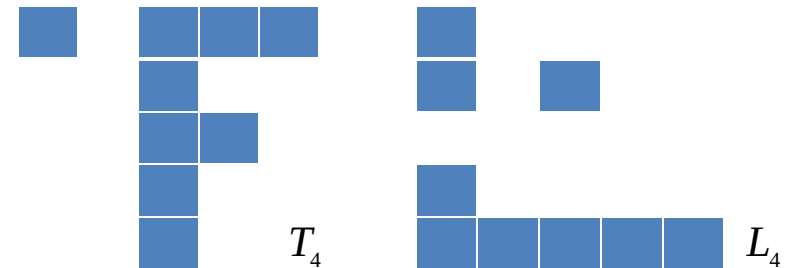
$$L_1 = [1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 1 \ 1]$$

Bu verilerin yanı sıra bozuk veriler de verelim....

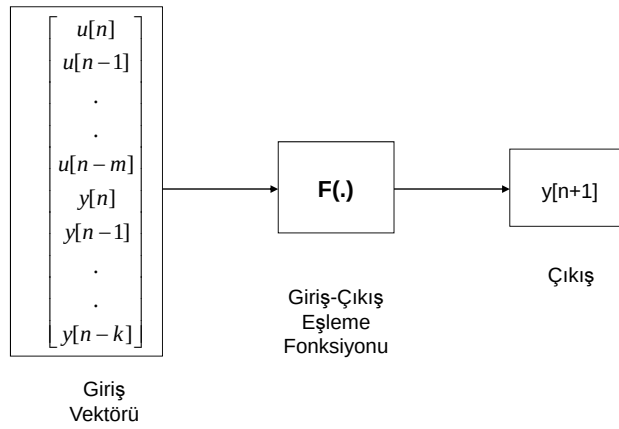


Bir de test kümesi oluşturalım...

Test kümesinde sağlam veriler ve eğitim kümesindekilerden farklı bozuk veriler olsun

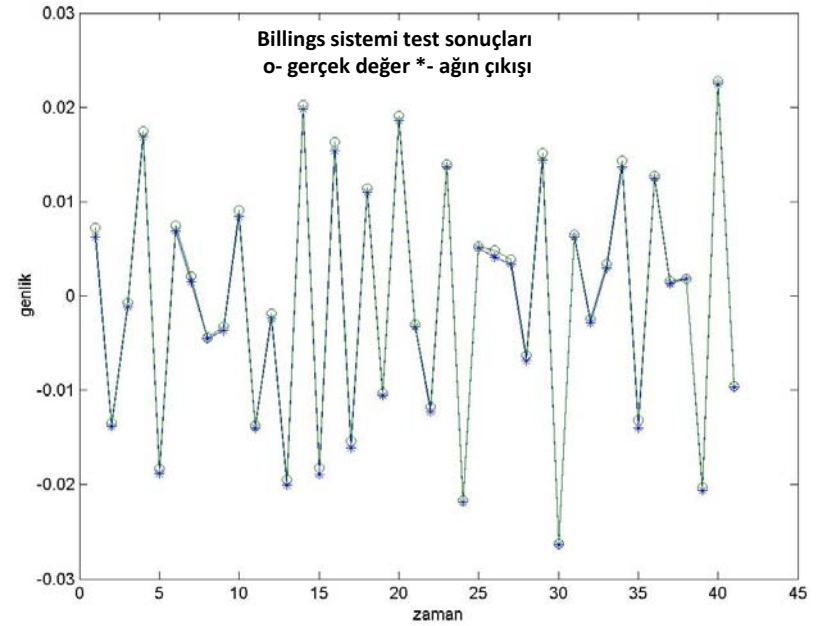
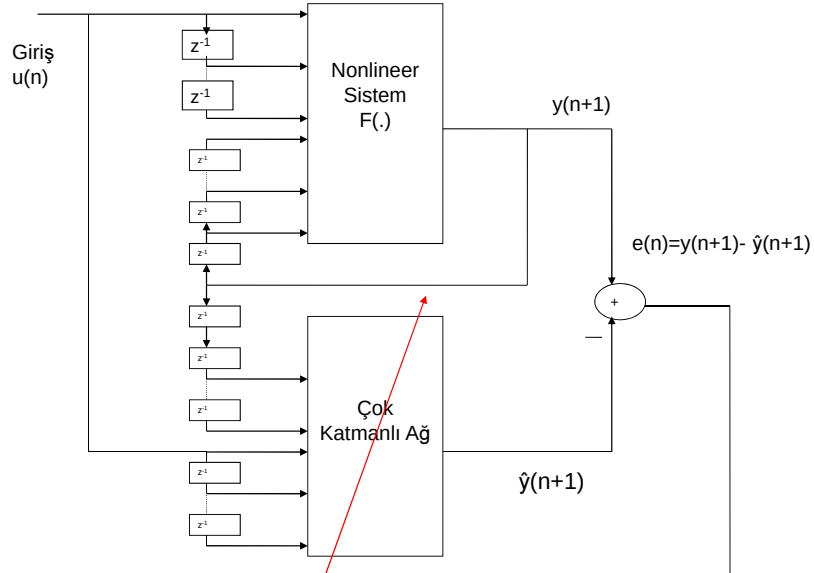
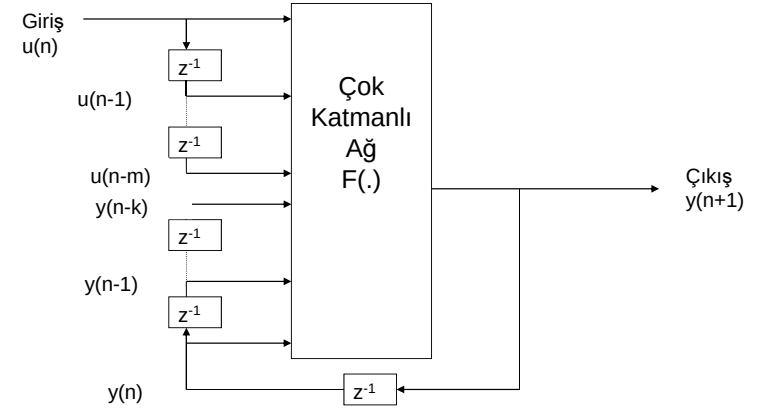


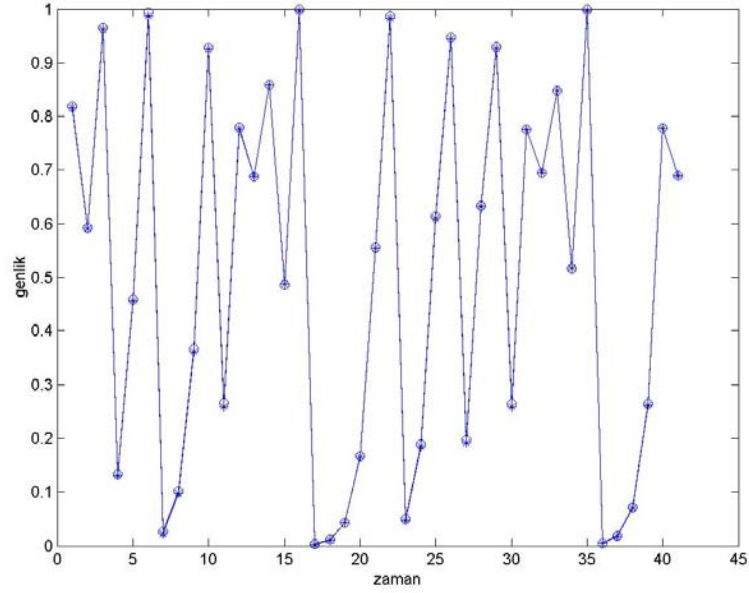
Giriş – Çıkış Modeline göre Dinamik Sistem Tanıma



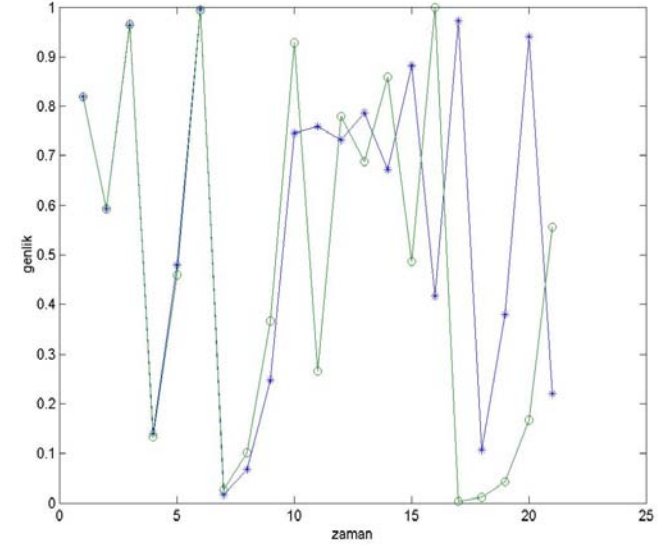
NARX (nonlinear autoregressive with exogenous inputs) modeli

$$y(n+1) = F(y(n), y(n-1), \dots, y(n-k), u(n), u(n-1), \dots, u(n-m))$$

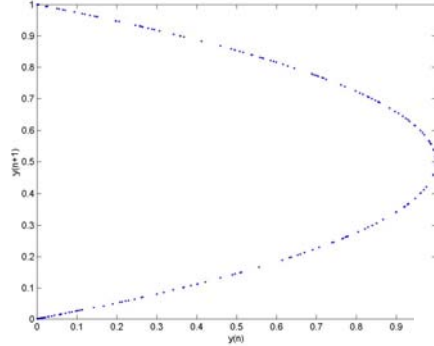




Feigenhoum sistemi için bir adım sonrasının öngörümü o- gerçek değer *- ağın çıkışı

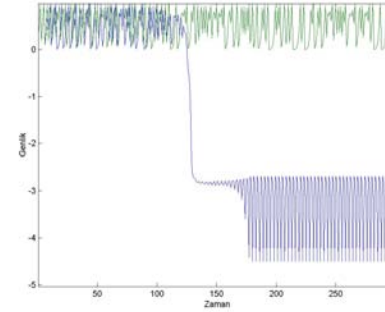
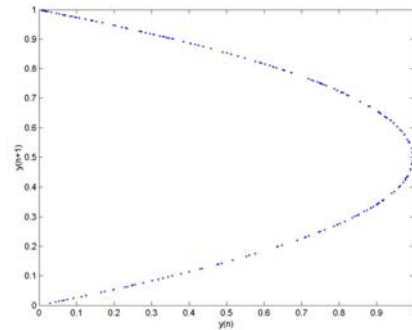


Feigenhoum sisteminin otonom davranışı o- gerçek değer *- ağın çıkışı



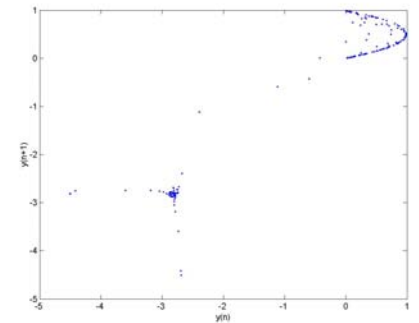
Gerçek sistemin çekicisi

Çok katmanlı ağın çekicisi



Çok katmanlı ağın çıkışı

Çok katmanlı ağın çekicisi



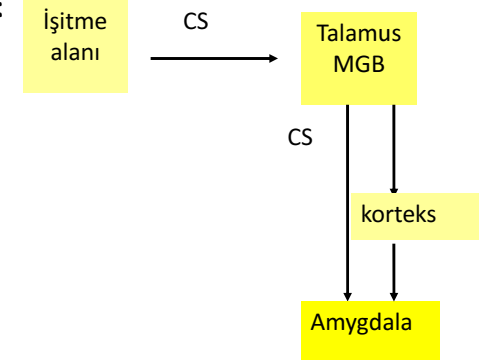
Bu yapı farklı bir öğrenme kuralı ile korku şartlanması için kullanılmıştır....

Armony, Servan-Schreiber, Cohen & LeDoux (1997)

- CS (Conditioned Stimulus) şartlı uyaran
- US (Unconditioned Stimulus) şartsız uyaran
- CS-US eşleştirilmesi tipik davranışlar fizyolojik tepkiler

Korku Şartlanması

Şartlanma sırasında:



Serkan Çapkan, Lisans Bitirme Ödevi, 2007

Yapılan Nedir?

Bir ses tonu ile birlikte, rahatsız edici bir etki uygulayarak beyin-deneğin-modelin o sese karşı şartlanmasını modellemek...

Yapılan Nedir?

Örnek;

Bir insana, çok sayıda farklı ses veriliyor.

Bu seslerin herhangi biri ile birlikte, bu insanın ayağına küçük bir elektrik şoku uygulanıyor.

Daha sonra yine sesler dinletilirken "o" ses verildiğinde, kişinin korkması bekleniyor

Yapılan Nedir?

Bu çalışmada;

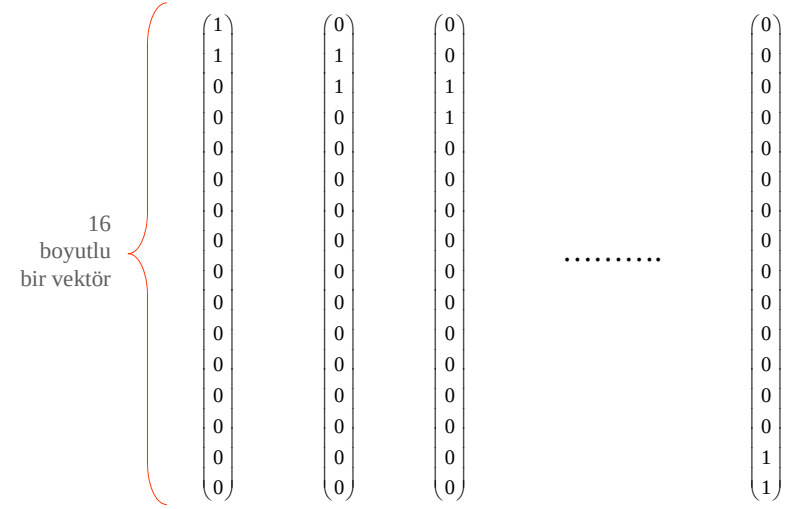
Modele,
15 farklı ses veriliyor, bu seslere tepkisi ölçülüyor.

Sonra,
bu seslerden herhangi biri ile birlikte, modele rahatsız edici bir uyarı uygulanıyor.

Daha sonra yine sesler aynı verilirken "o ses" verildiğinde, modelin beklenmeyen bir tepki vermesi bekleniyor

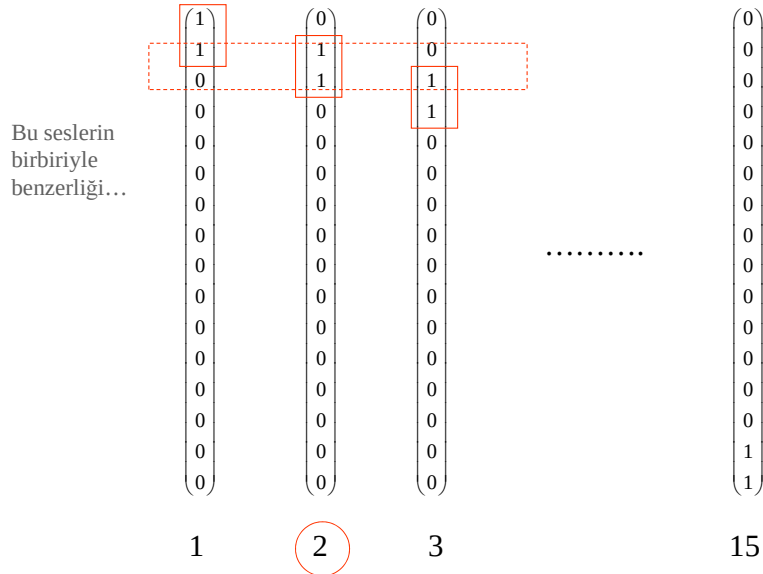
Ses Nasıl Modelleniyor?

(15 farklı frekansta ses)



Ses Nasıl Modelleniyor?

(15 farklı frekansta ses)



İşlem Basamakları

Seslerin Uygulanması, tepkilerin ölçülmesi

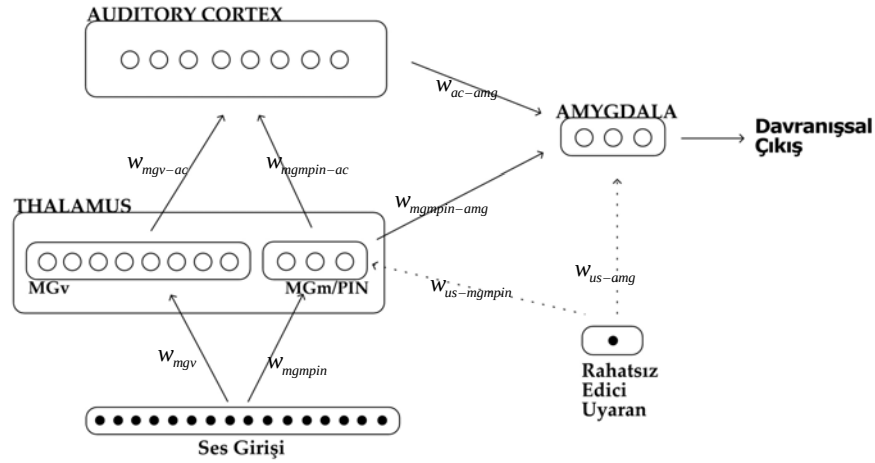


Yine seslerin uygulanması, 5. seste rahatsız edici uyarının uygulanması

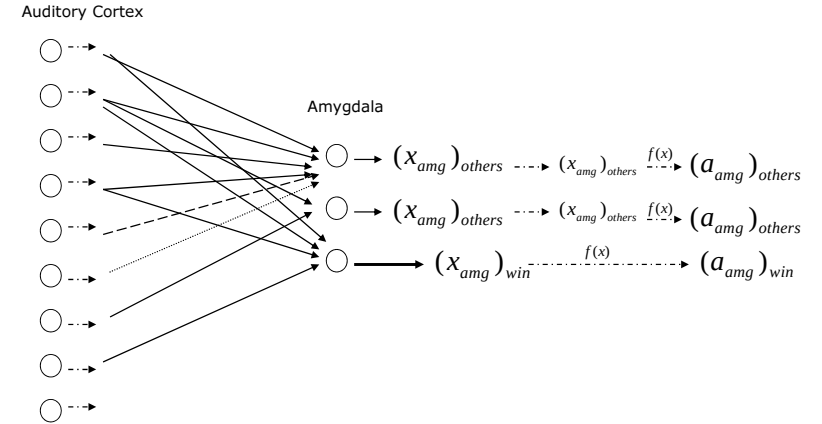


Gözlem

Model



Örnek Bir İşlem (auditory cortex -> amygdala için)

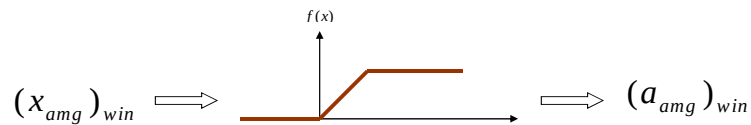


Hücreye gelen girişin ağırlıklandırılması

$$(w_{ac-amg}) * (a_{ac}) = (x_{amg})$$

Örnek Bir İşlem (auditory cortex -> amygdala için)

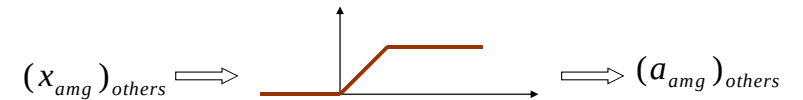
Kazanan hücre çıkışının aktivasyon fonksiyonunda geçirilmesi



Örnek Bir İşlem (auditory cortex -> amygdala için)

Diğer hücre çıkışlarının aktivasyon fonksiyonundan geçirilmesi

$$(x_{amg})_{others} = (x_{amg})_{others} - \mu * (a_{amg})_{win}$$



Tüm aktive edilmiş hücre çıkışları

$$a_{amg} = (a_{amg})_{win} + (a_{amg})_{others}$$

Örnek Bir İşlem (auditory cortex -> amygdala için)

Ağırlıkların Yenilenmesi

Yalnızca değeri o bölgenin çıkışlarının ortalamasından büyük olan çıkışlara ilişkin ağırlıklar yenilenir

$$w_{rs}' = w_{rs} + \epsilon . a_r . a_s , \quad a_s > a_{ort}$$

$$w_{rs}' = w_{rs} , \quad \text{diğerleri}$$

İşlem Basamakları

Seslerin Uygulanması



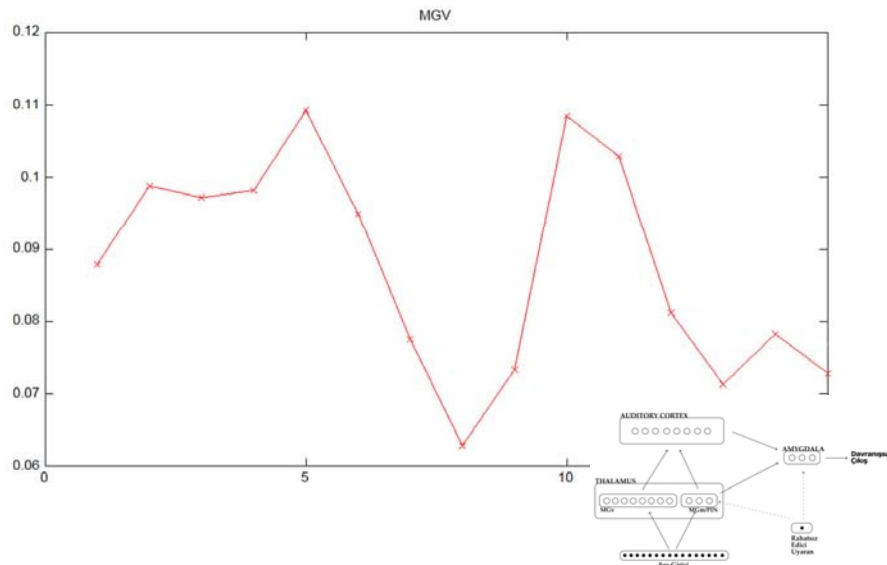
Sesler ile birlikte rahatsız edici uyarının uygulanması



Gözlem

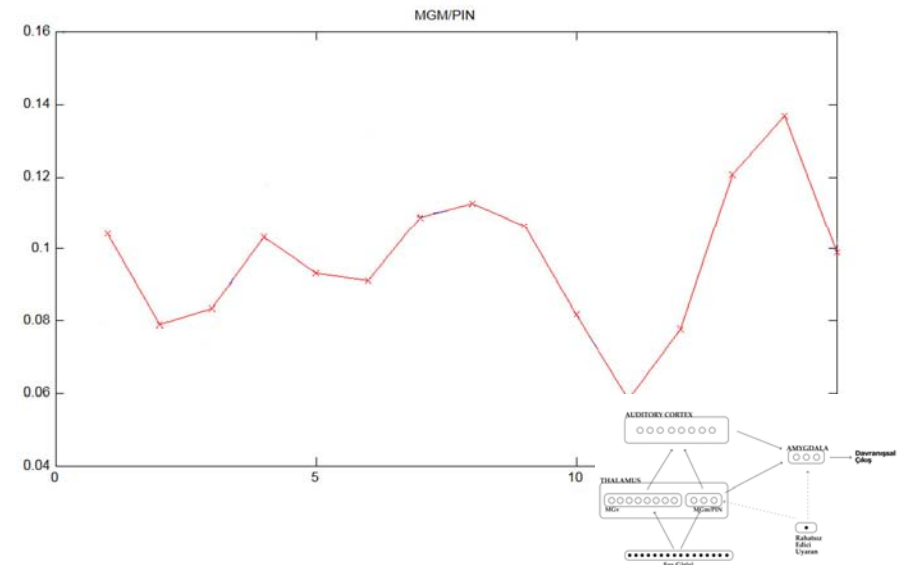
Sonuçların İncelenmesi

(Seslerin İlk Uygulanması)



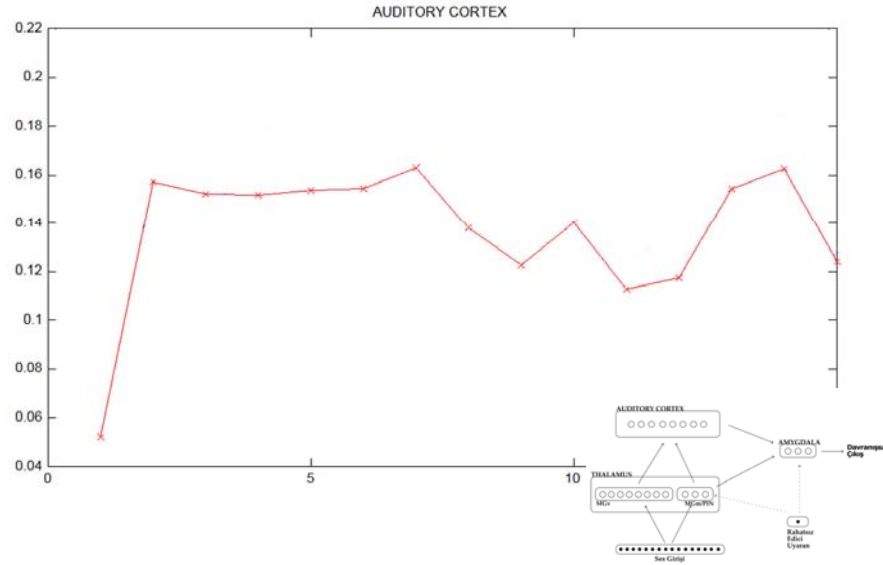
Sonuçların İncelenmesi

(Seslerin İlk Uygulanması)



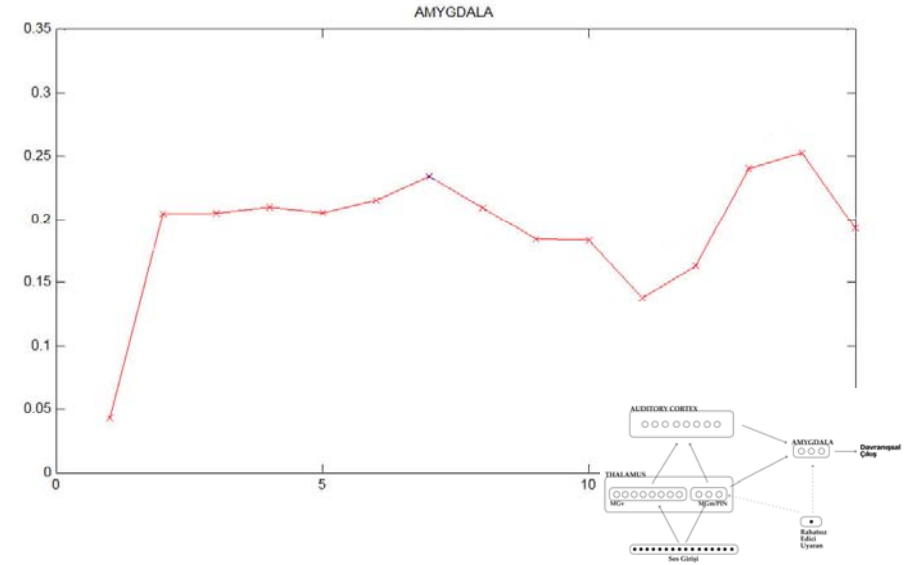
Sonuçların İncelenmesi

(Seslerin İlk Uygulanması)



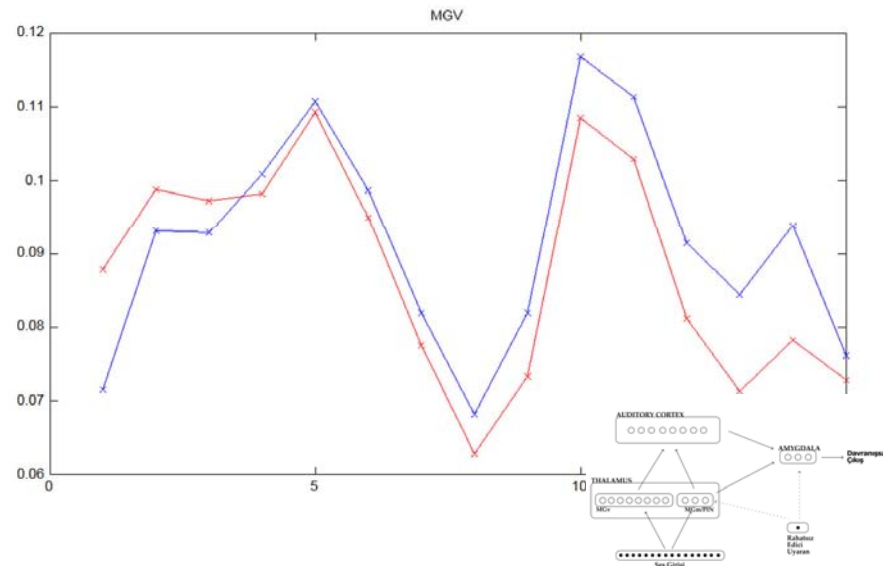
Sonuçların İncelenmesi

(Seslerin İlk Uygulanması)



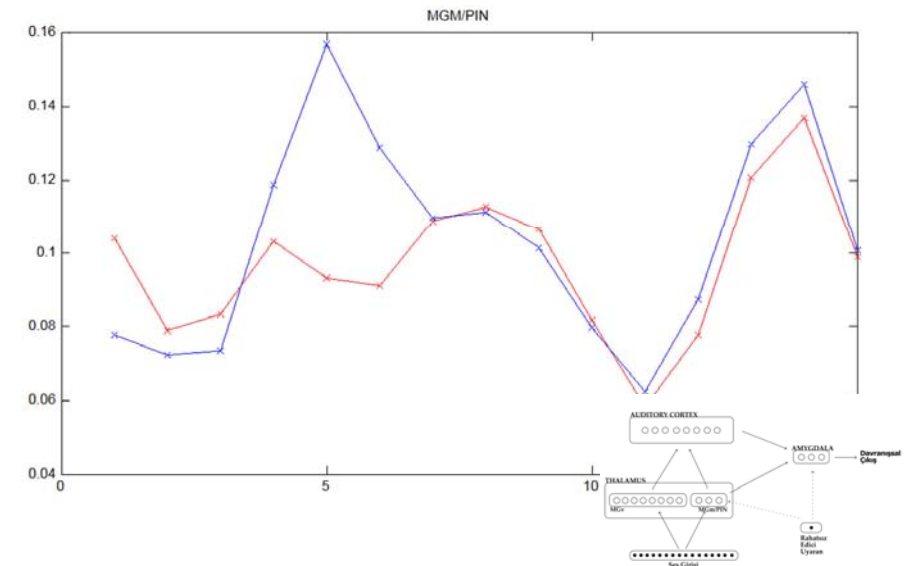
Sonuçların İncelenmesi

(Şartlanmadan sonra ve öncekiler)



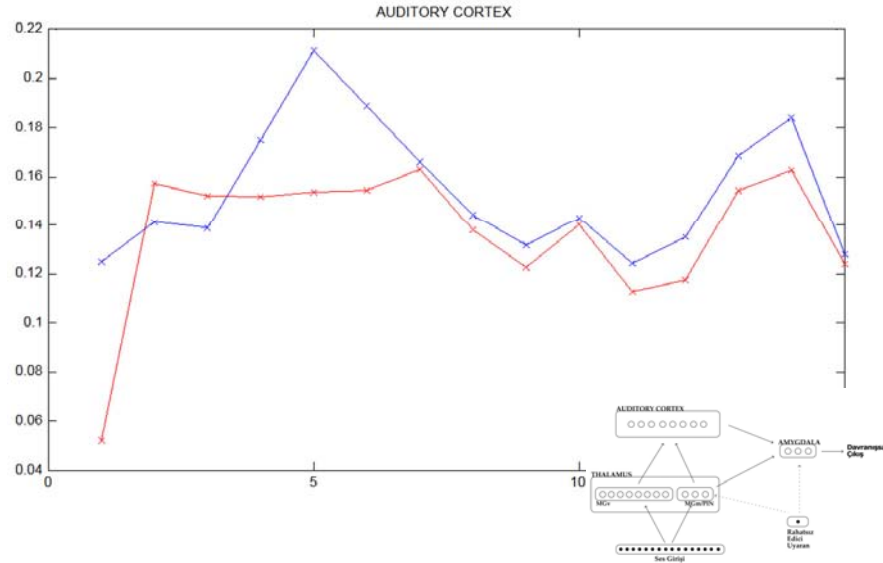
Sonuçların İncelenmesi

(Şartlanmadan sonra ve öncekiler)



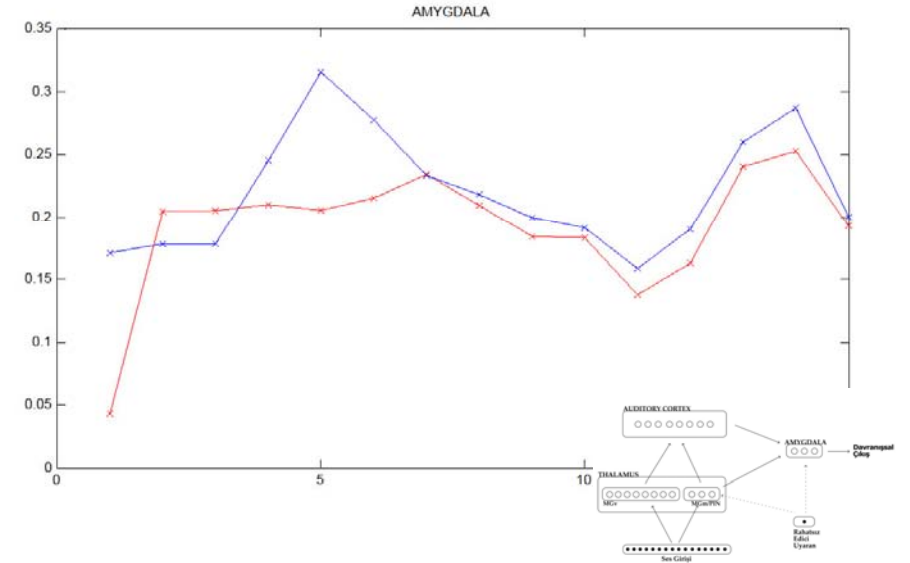
Sonuçların İncelenmesi

(Şartlanmadan sonra ve öncekiler)



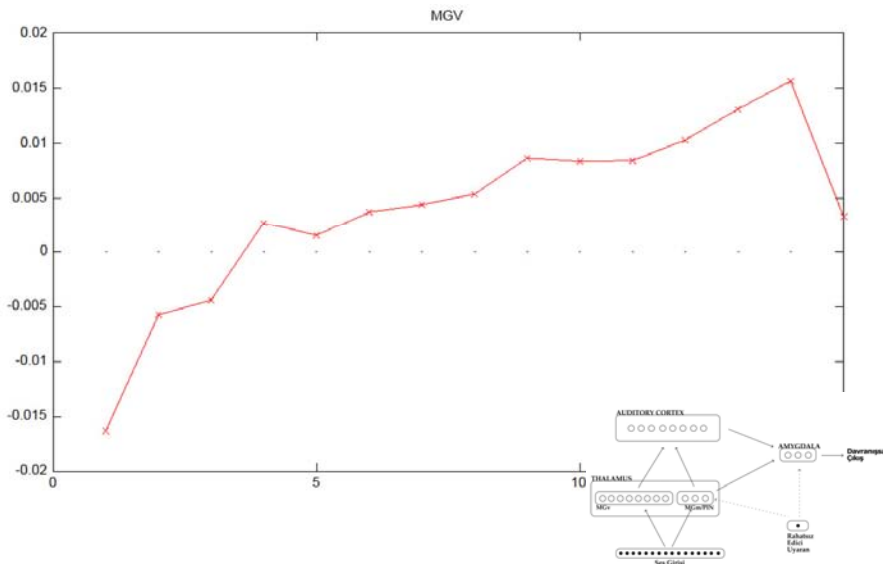
Sonuçların İncelenmesi

(Şartlanmadan sonra ve öncekiler)



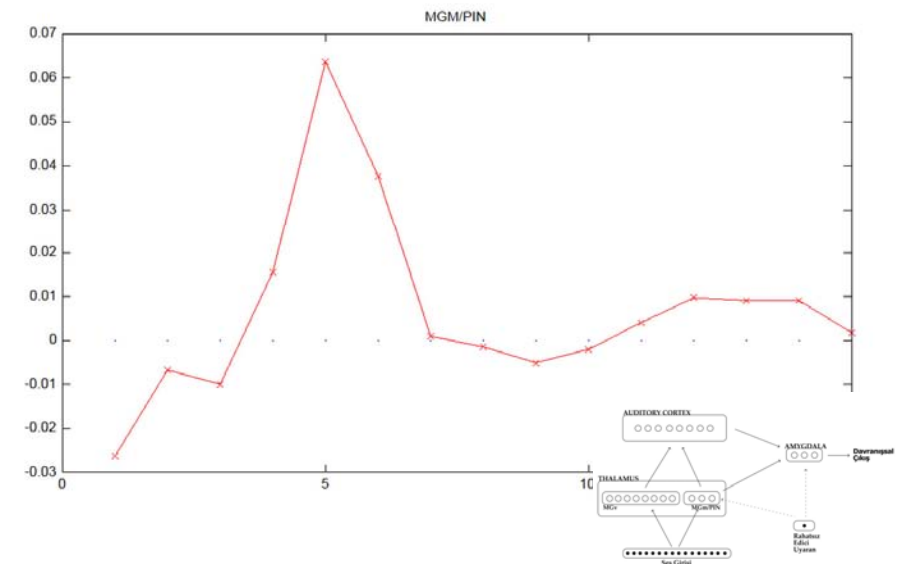
Sonuçların İncel.

(Şartlanma öncesi ve sonrası çıkışların farkları)

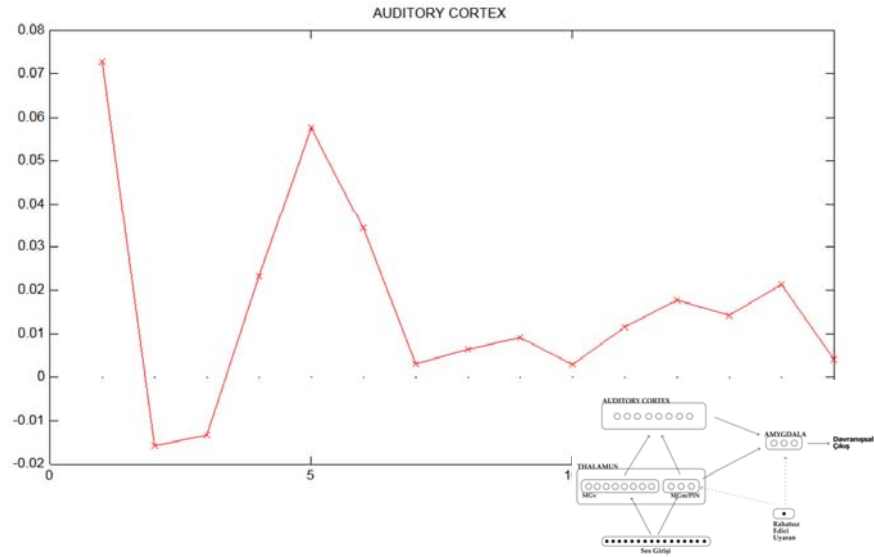


Sonuçların İncel.

(Şartlanma öncesi ve sonrası çıkışların farkları)



Sonuçların İncel. (Şartlanma öncesi ve sonrası çıkışların farkları)



Sonuçların İncel. (Şartlanma öncesi ve sonrası çıkışların farkları)

