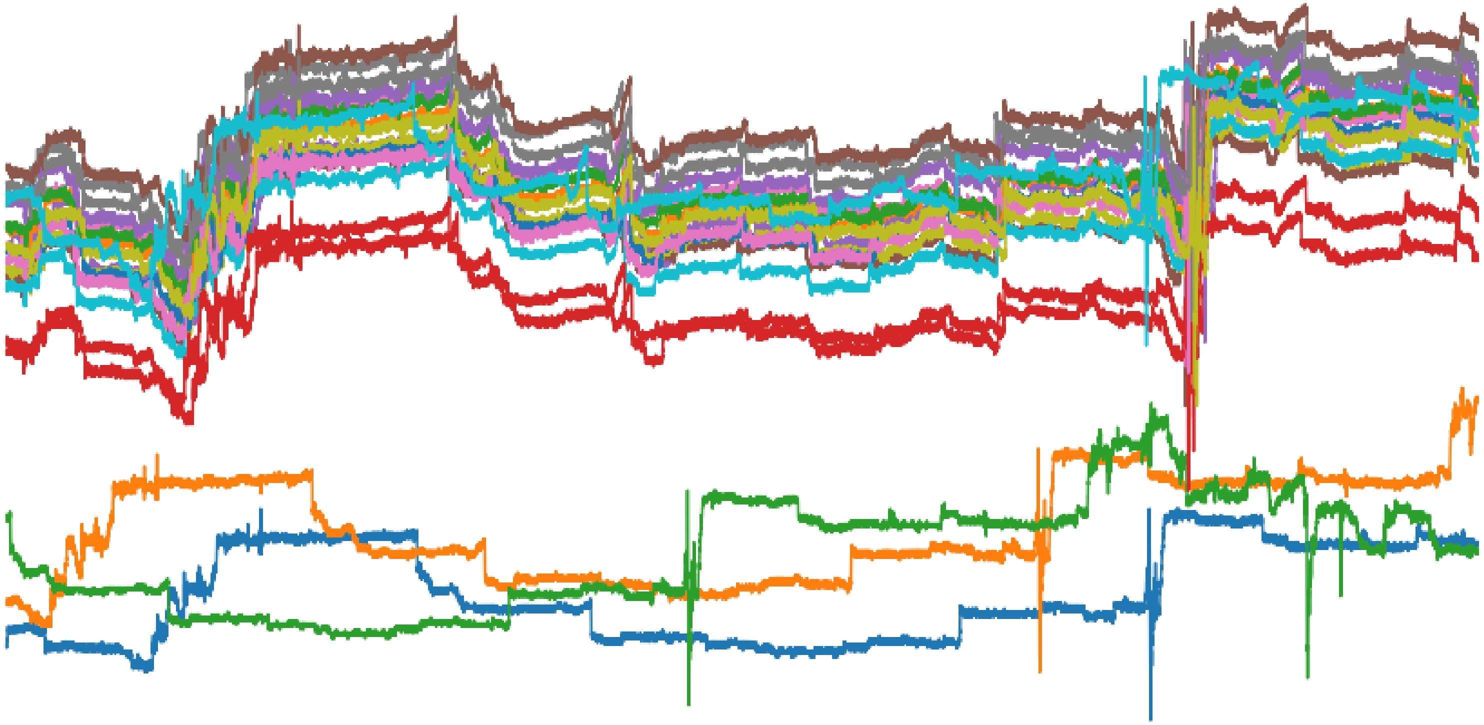
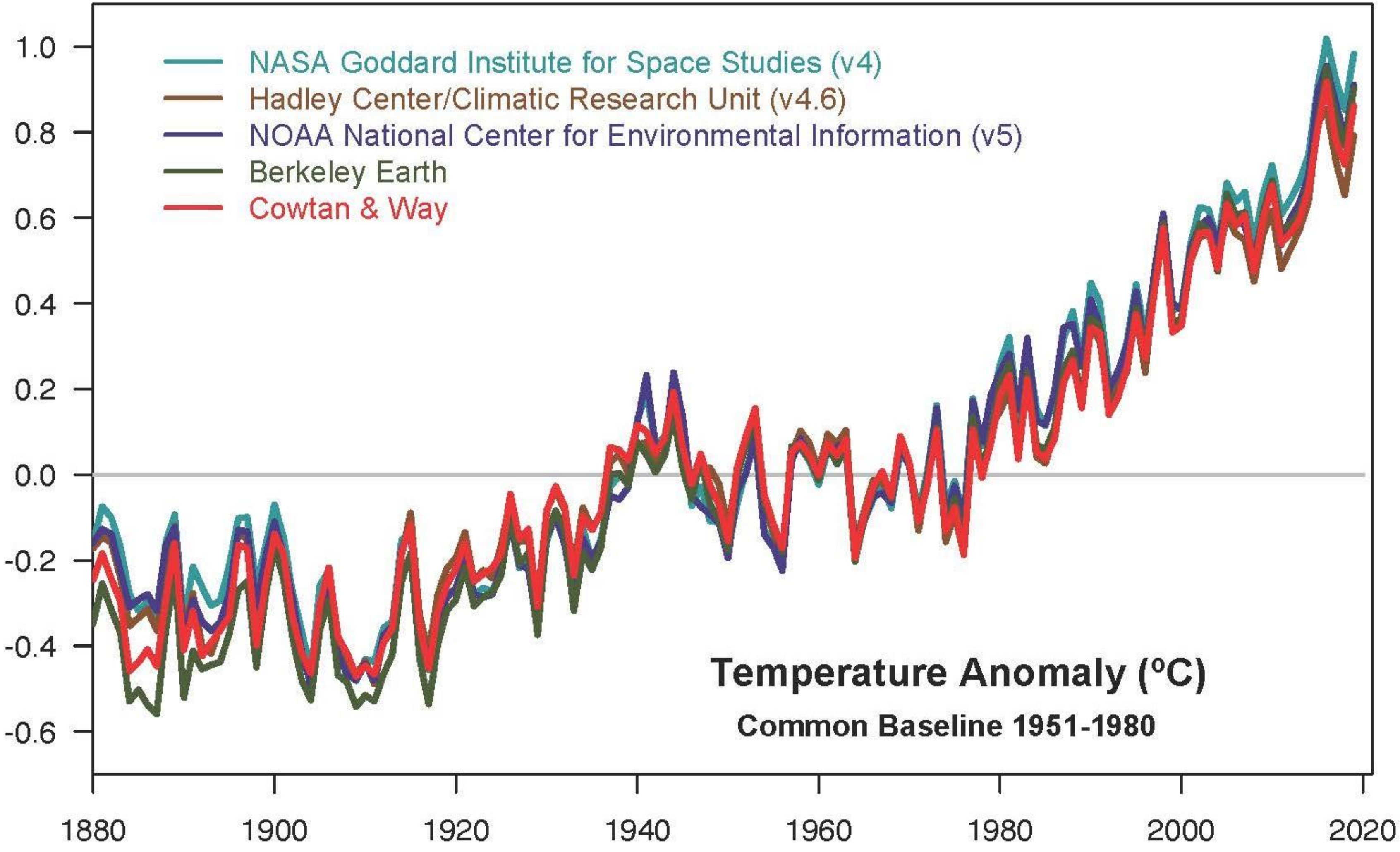
Çok Boyutlu Zaman Serilerinde Boyut Küçültme



Dr. Öğr. Üyesi Barış Akgün

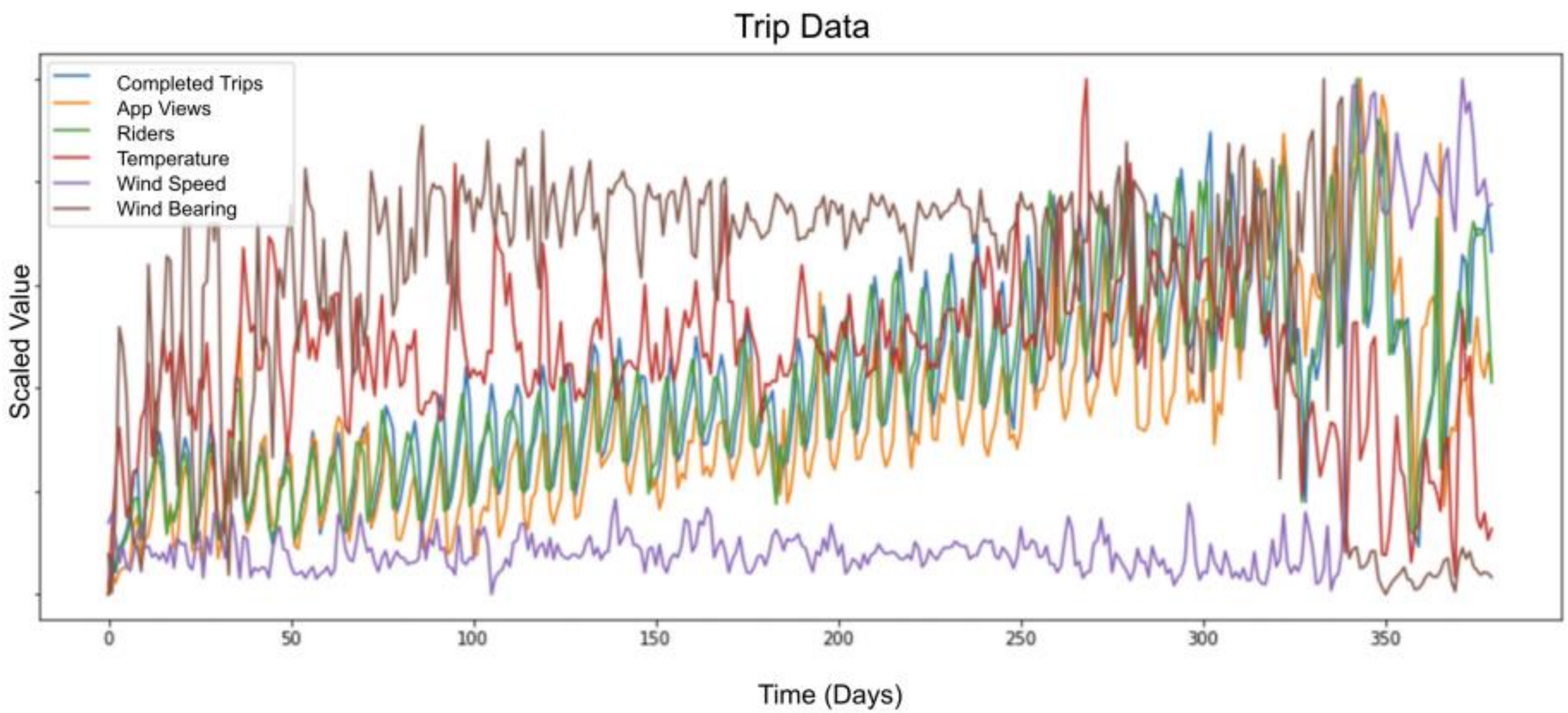
Koç University

Küresel Isınma



Piyasalar

Uber Araç Kullanımı



Zaman Serileri Her Yerde

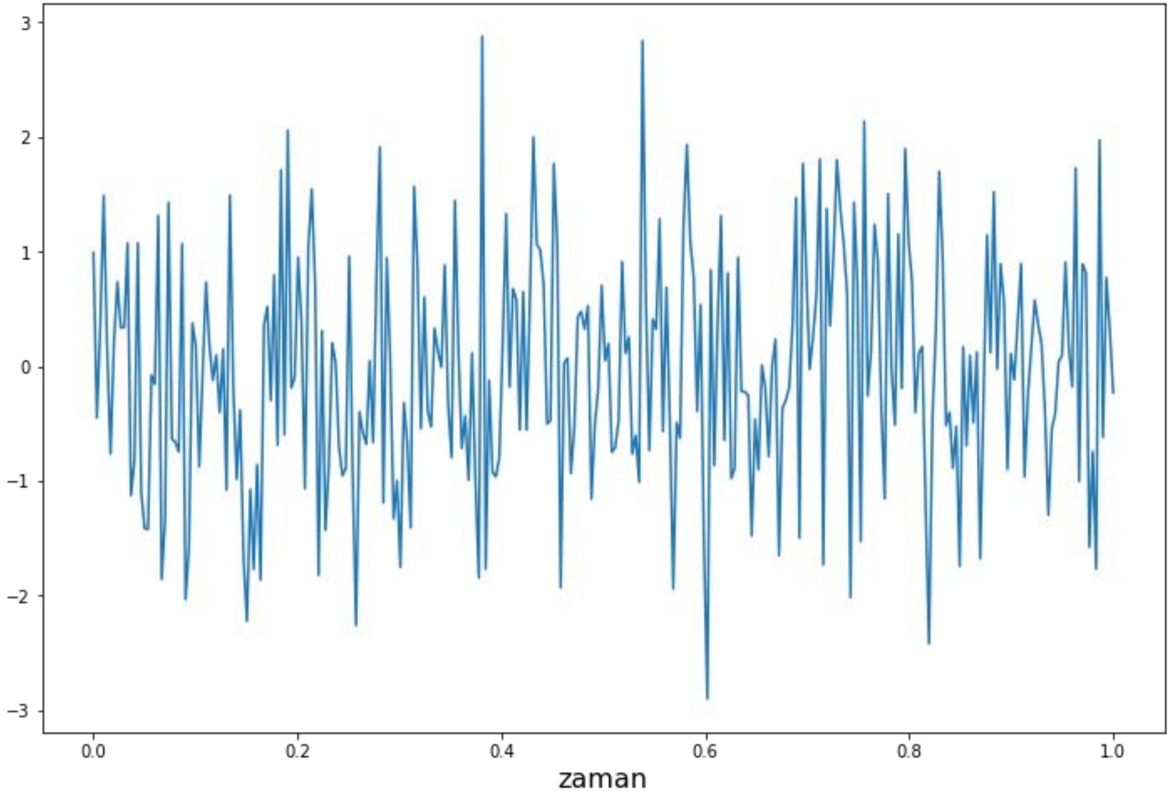
- Dijitalleşmiş fabrikalar
- Kentler
- Akıllı telefonlar
- Uydular
- Bilimsel veri toplama araştırmaları
- Ses, Video
- Robot/Drone/Otonom Araç Sensörleri
- Metin ve Gen Verileri (Tam olarak zaman serisi değil ama sıralı yinede sıralı)
- …

Zaman Serileriyle Ne Yapmak İsteriz?

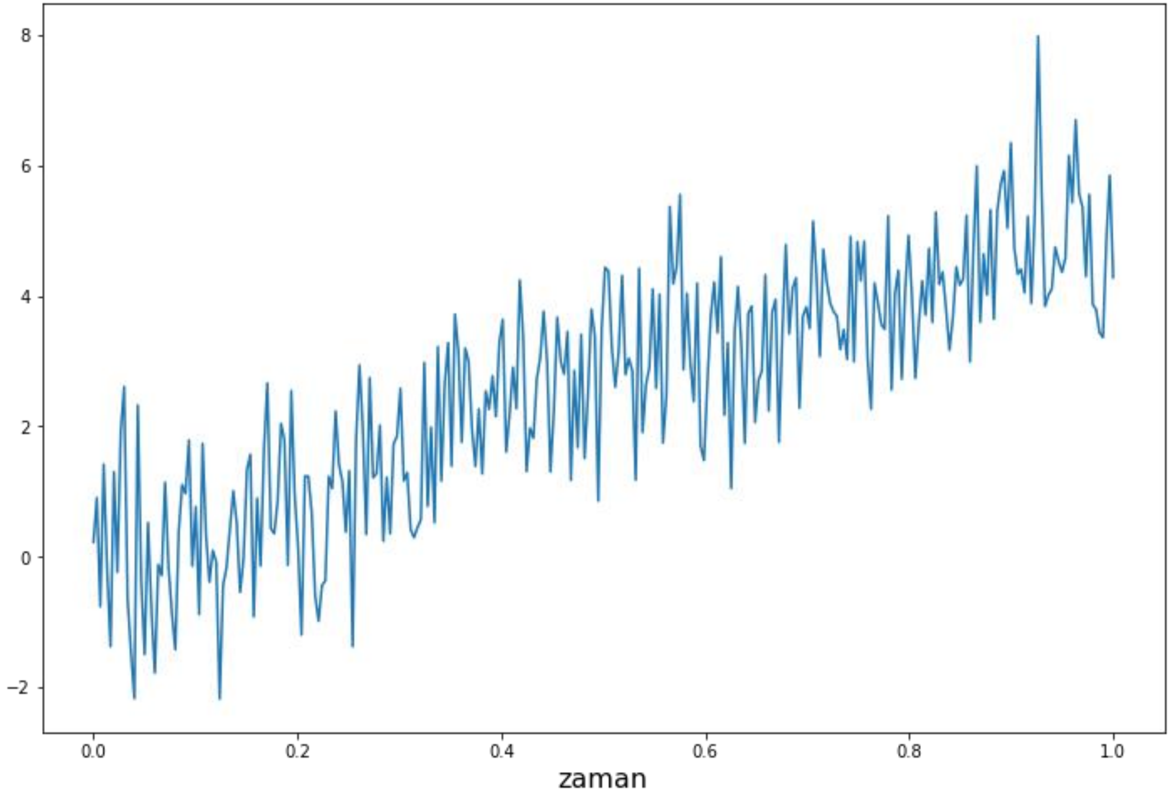
- Gelecek değerlerini tahmin etmek (forecasting)
  - Örn: Yarın borsa ne olacak?
- Gözlemlenen bir serinin sınıflandırılması
  - Örn: Akıllı saat verilerinden aktivite tanıma
- Aykırılık (anomaly) Tanımlama
  - Örn: Rafineri süreci sıcaklık ölçümleri
- Kümeleme
  - Örn: Farklı kuş şarkılarını gruplama
- Arama/Endeksleme
  - Örn: Dinlediğim şarkının adını ya da dinlediğim şarkıya en benzer şarkıyı bulma
- Değişim zamanı bulma
  - Makroekonomik trendlerin değişme zamanları

Sadece bunlarla sınırlı değil!

Temel Zaman Serisi Tipleri

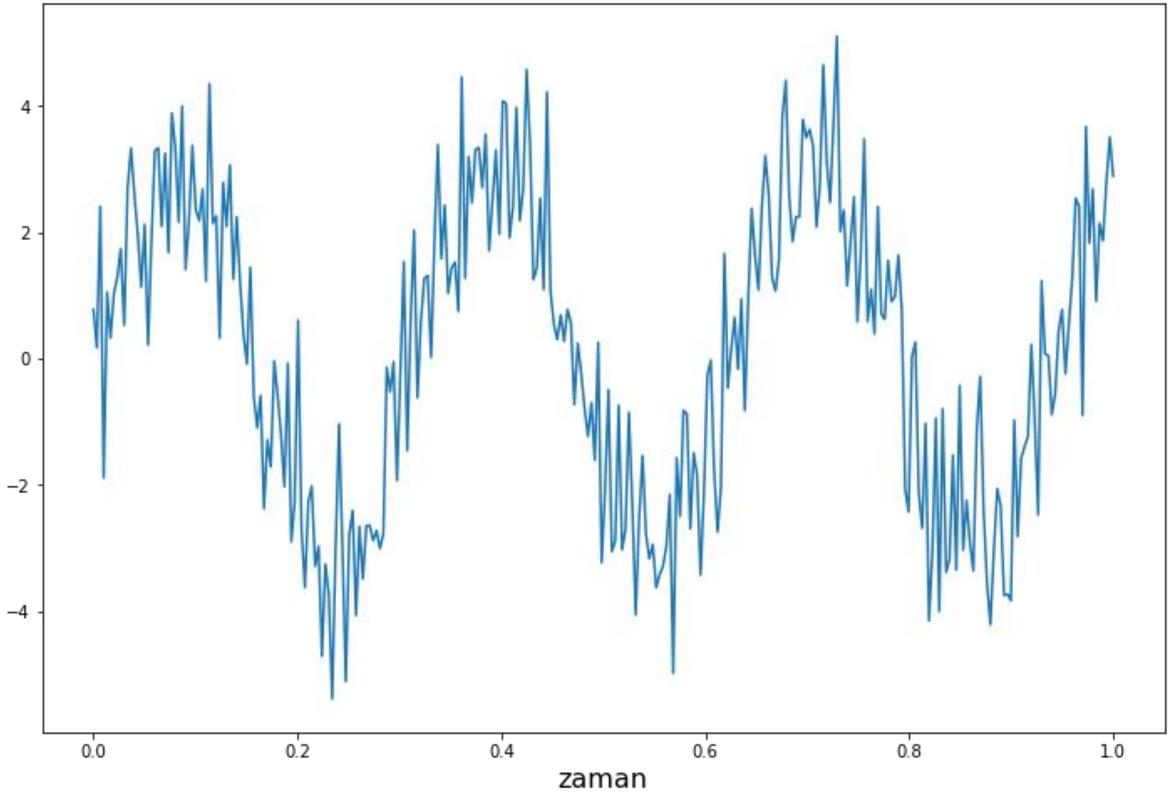


Durağan

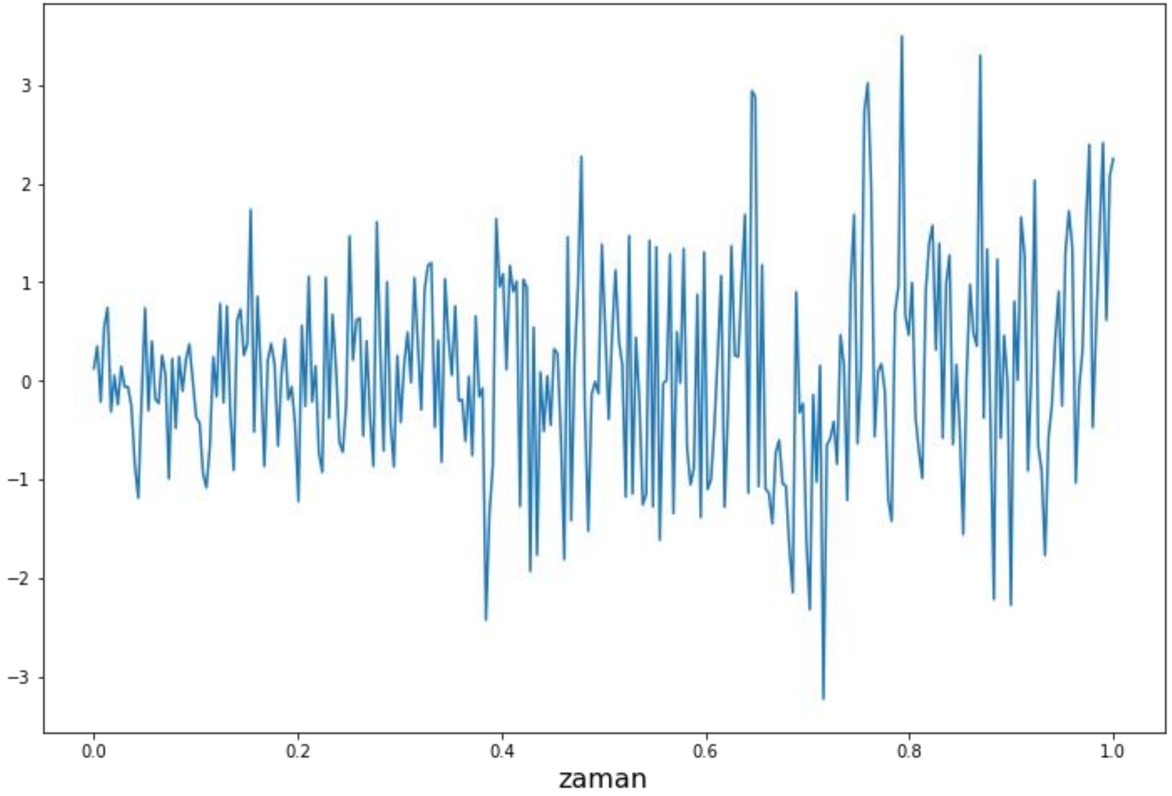


Trend

Temel Zaman Serisi Tipleri



Sezonsal



Artan Varyans

- Bir zaman serisi gördüklerimizin karışımından da oluşabilir
- Daha karmaşık da olabilir

Gecikme

- Elimizdeki zaman serisi:

$$X = x_1, x_2, \dots, x_{t-1}, x_t, x_{t+1}, \dots$$

- x_t'ye göre i'inci gecikme

$$G(x_t, i) = x_{t-i}$$

- Birinci derece sezonsal olmayan fark:

$$\Delta x_t = x_t - x_{t-1}$$

- İkinci derece sezonsal olmayan fark (farkların farkı)

$$\Delta^2 x_t = \Delta x_t - \Delta x_{t-1} = x_t - 2x_{t-1} + x_{t-2}$$

- Sezonsal fark (sezon periyodu s)

$$\Delta_s x_t = x_t - x_{t-s}$$

Otokorelasyon

μ: Ortalama
σ: Standard sapma

- Kovaryans:

$$cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_X)(Y_i - \mu_Y)$$

- Korelasyon:

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_y}$$

Otokorelasyon

μ: Ortalama
σ: Standard sapma

- Otokorelasyon

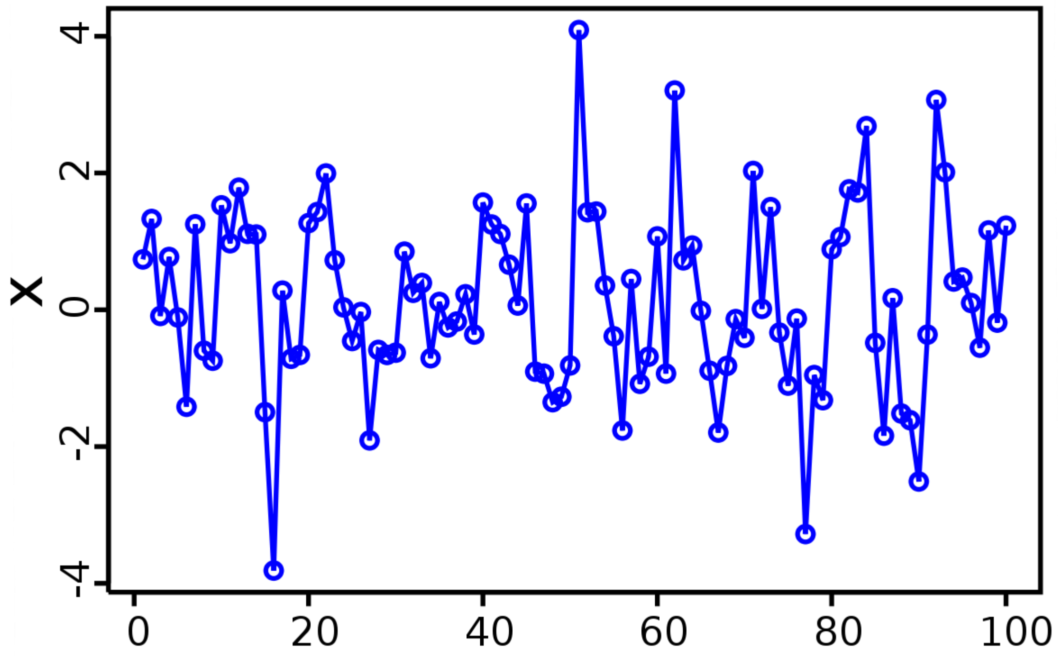
$$autocorr(X, i) = \frac{cov(X, G(X, i))}{\sigma_X \sigma_{G(X,i)}}$$

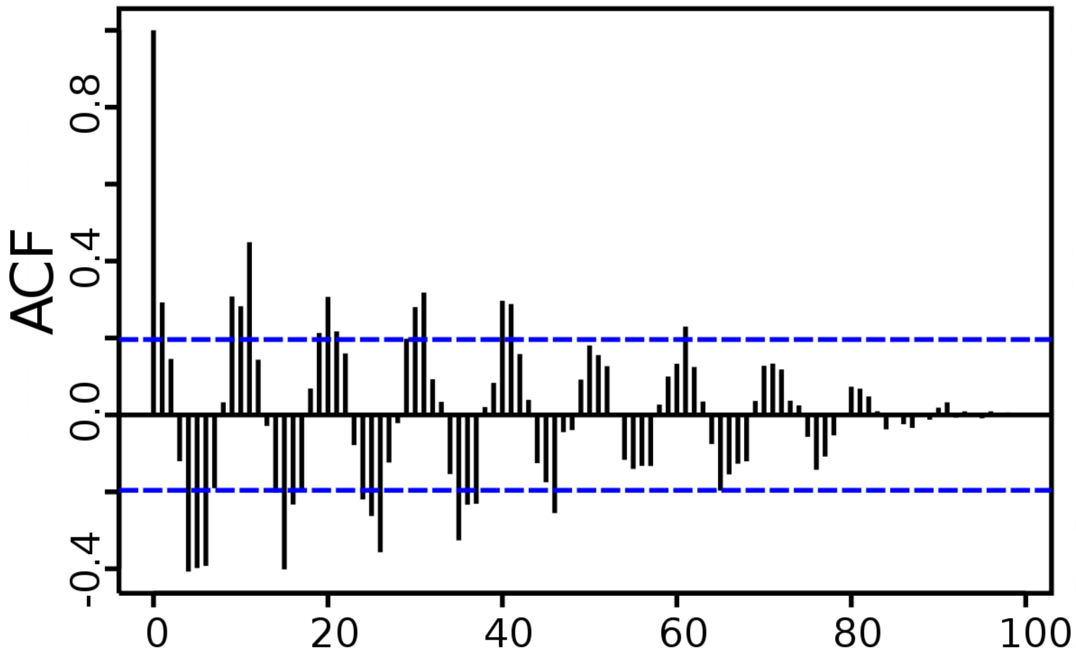
- Durağanlık varsayımı altında $\sigma_X = \sigma_{G(X,i)}$:

$$autocorr(X, i) = \frac{cov(X, G(X, i))}{\sigma_X^2}$$

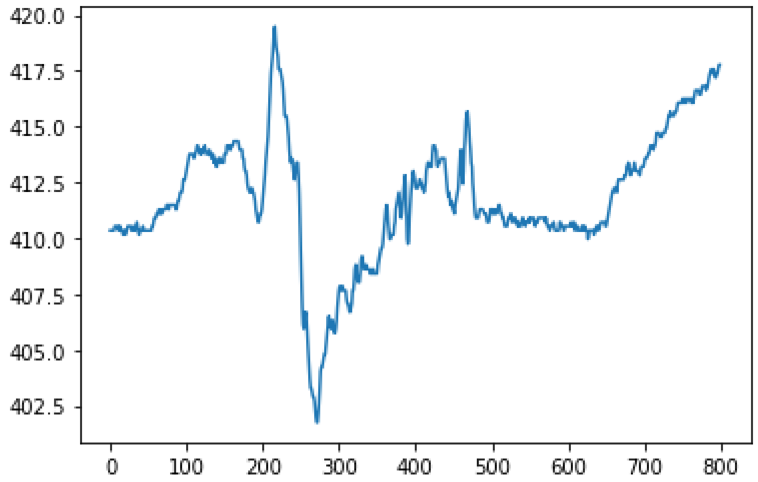
- Kısmi Otokorelasyon: Daha kısa gecikme etkilerinden arındırılmış otokorelasyon

Wikipedia'dan Örnek:

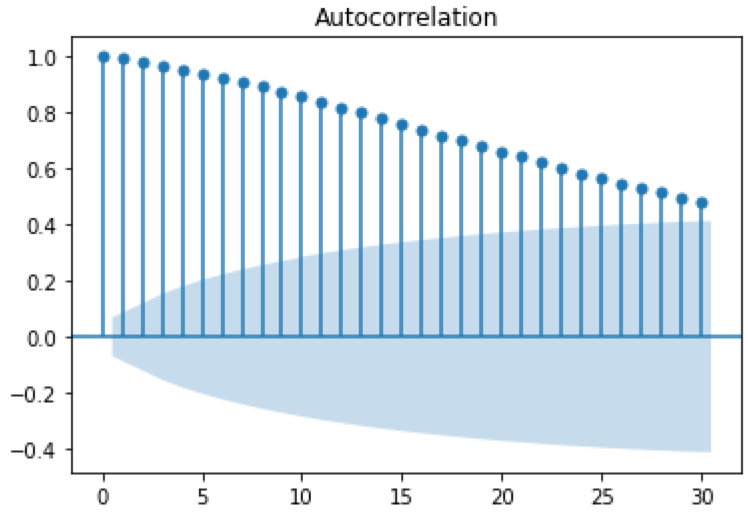




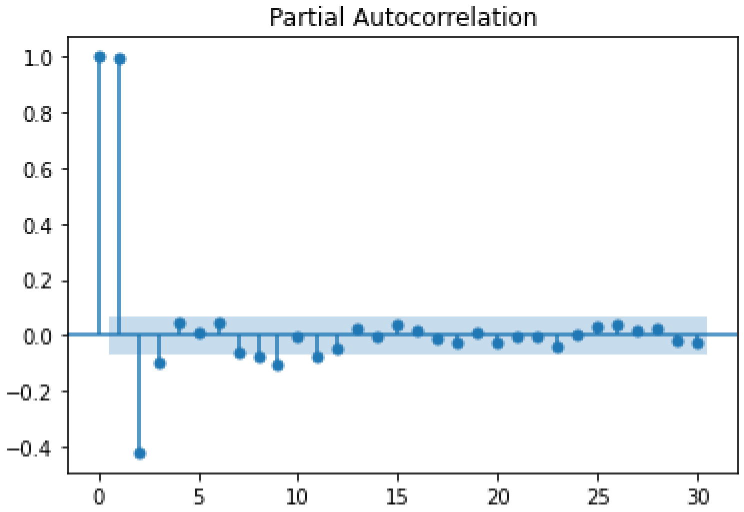
Örnek



t



gecikme



gecikme

Bazı Modeller

Ak gürültü modeli: $\mathcal{N}(0, \Sigma_\epsilon)$

- Regresyon modelleri

$$y_t = w^T X_t + \epsilon_t$$

  - y: scalar çıktı, X: d boyutlu değişken, w: parametre vektörü, ϵ: scalar ak gürültü

- Otoregresyon modelleri

$$X_t = A X_{t-1} + E_t$$

  - A: $d \times d$ kare parametre matrisi, E: ak gürültü vektörü

Bazı Modeller

- Hareketli Ortalama Modelleri

$$X_t = AX_t + \Theta E_{t-1} + E_t$$

  - Θ: $d \times d$ hareketli ortalama parametre matrisi, E_{t-1}: tamin edilen X_{t-1} ile gerçek X_{t-1} arasındaki fark (ak gürültü olarak varsayılır)

- Kırılma noktalı modeller

$$X_t = A_1 X_{t-1} I(t \leq \tau) + A_2 X_{t-1} I(t > \tau)$$

  - τ: kırılma noktası, I: indikatör fonksiyonu (içi doğru ise 1, değilse 0)

Bazı Modeller

- Kontrol modelleri

$$X_T = AX_{t-1} + BU_t + E_t$$

  - B: $d \times d$ kare matris, U: d –boyutlu control değişkeni

- Faktör modelleri

$$X_t = \Lambda F_t + E_t$$

  - F: k boyutlu faktörler/gizli değişkenler ($k < d$), Λ: $\mathrm{d} \times k$ parametre matrisi
  - Faktörler de dinamik olabilir:

$$F_t = \Phi F_{t-1} + U_t$$

ARIMA Modelleri

- ARIMA(p,d,q):
  - p: Otoregresyon terim sayısı
  - d: Sezonsal olmayan fark derecesi
  - q: hareketli ortalama terim sayısı
  - (Ufak Not: p ve d çok alakalı)

- Genel model (p>d için):

$$x_t = \mu + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + \epsilon_t$$

- x_t skalar

Klasik Yaklaşım

1. Seriyi durağan bir hale getir
   - Ardışık değerlerin farkını al (trend için)
   - Sezonsal değerlerin farkını al (sezonsallık için)
   - Logaritmasını al ya da indir (deflate)
2. Otokorelasyon ve kısmi otokorelasyon değerlerine bakıp, ARIMA parametrelerini seç
   - Bu konuda yardımcı olacak kaynaklar mevcut
3. Tahmin ile gerçek değerlerin farkına bak:
   - Ak gürültü gibi davranıyorsa işimiz bitti
   - Değilse farkın otokorelasyon ve kısmi otokorelasyon değerlerini incele ve 2. adıma dön

Çok Boyutlu: Vektör Otoregresyon Modeli (VAR)

- d boyutlu bir zaman serisi için:

$$X_t = \mu_X + A_1 X_{t-1} + A_2 X_{t-2} + \cdots + A_p X_{t-p} + \epsilon_t$$

  - $\epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon)$ (özdeş ve bağımsız olarka dağılmış, i.i.d)
  - A_i: $d \times d$ matrisler
- Örnek:
  - Varsayımlar:
    - BIST100 hisseleri birbirlerini etkiliyor (100 hisse)
    - 1 dakika arayla veri topluyoruz
    - Son 10 dakikaya bakacağız
  - Kaç tane parametre var?
    - $100 + 100 \times 100 \times 10 \approx 10^5$!!!

VAR

- A_i: matrislerinin seyrek (sparse) olduğu varsayımı altında Lasso yöntemi (parametrelerin l_1 düzenlileştirilmesi) ve onun varyantları ile çözülür
- Ancak yüksek boyutlu verilerde, Lasso yönteminin iyi çalışacağı kesin değildir
- Ne yapılabilir?
  - Problem hakkında öncül bilgiler kullanmak (örneğin tarım hisseleri ile teknoloji hisselerini bağımsız kabul etmek, tarım teknolojisi şirketleri hariç)
  - Daha az parametreli modeller denemek
  - Zaman serisinin boyutlarını küçültmek
  - Başka?

Az Paremetreli Modele Basit Örnekler

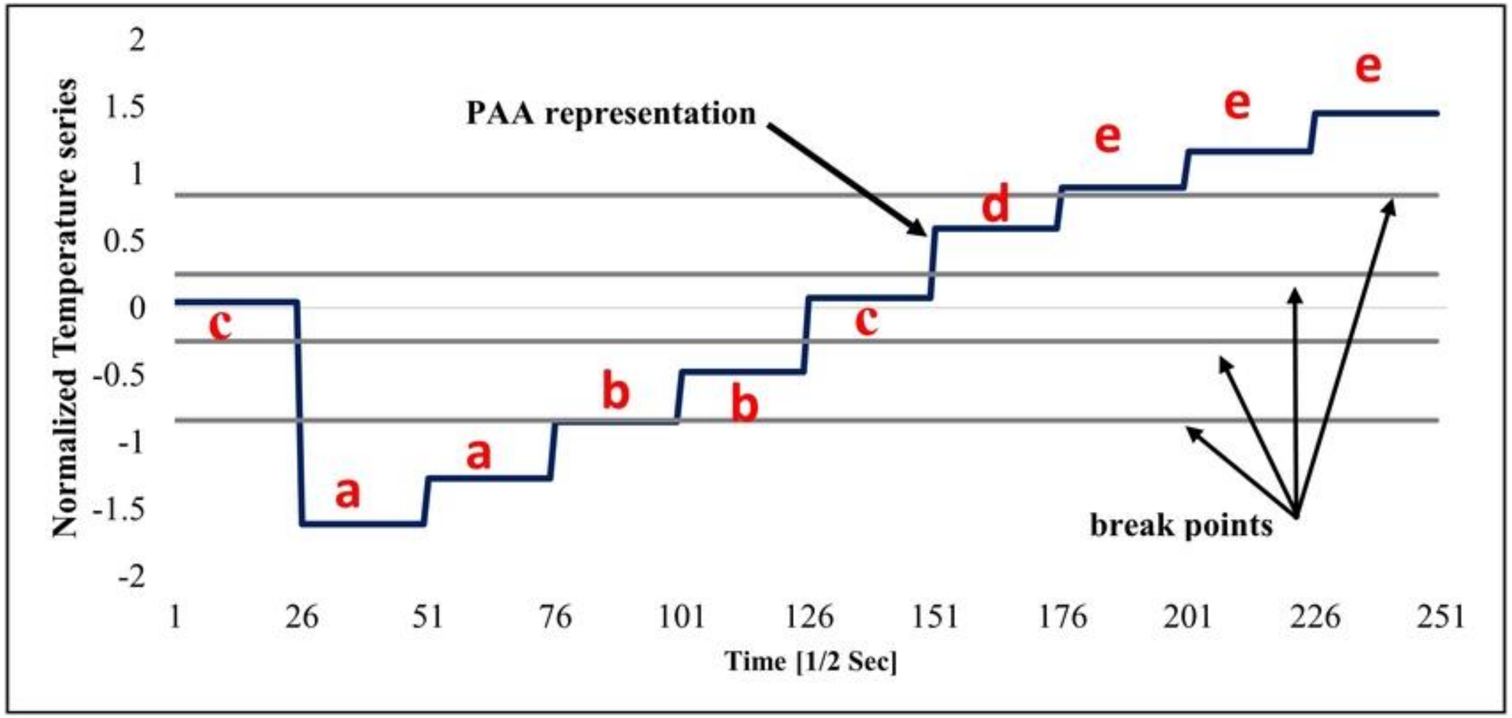
- Parametre matrislerini scalar değerlerle değiştirmek
$$X_t = \mu_X + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \epsilon_t$$
  - Hisse Örneği: 110 parametre
- Parametre matrislerini köşegen olarak varsayma
  - Hisse Örneği: 1100 parametre
- Faktör Modelleri kullanma?

Zaman Serilerinde Boyut Küçültmek

- Zamanda boyut küçültmek
  - Bazen zamandan bağımsız bir gösterim bulmak
- Ölçüm uzayında boyut küçültmek
- Hem zamanda hem ölçüm uzayında boyut küçültmek
- Sürekli ölçümlerden sembol çıkartmak
  - Hem zaman hem de ölçüm uzayında

Örnek: PAA ve SAX

- “Piecewise Aggregate Approximation” ve “Symbolic Aggregate Approximation”



Kabaca:

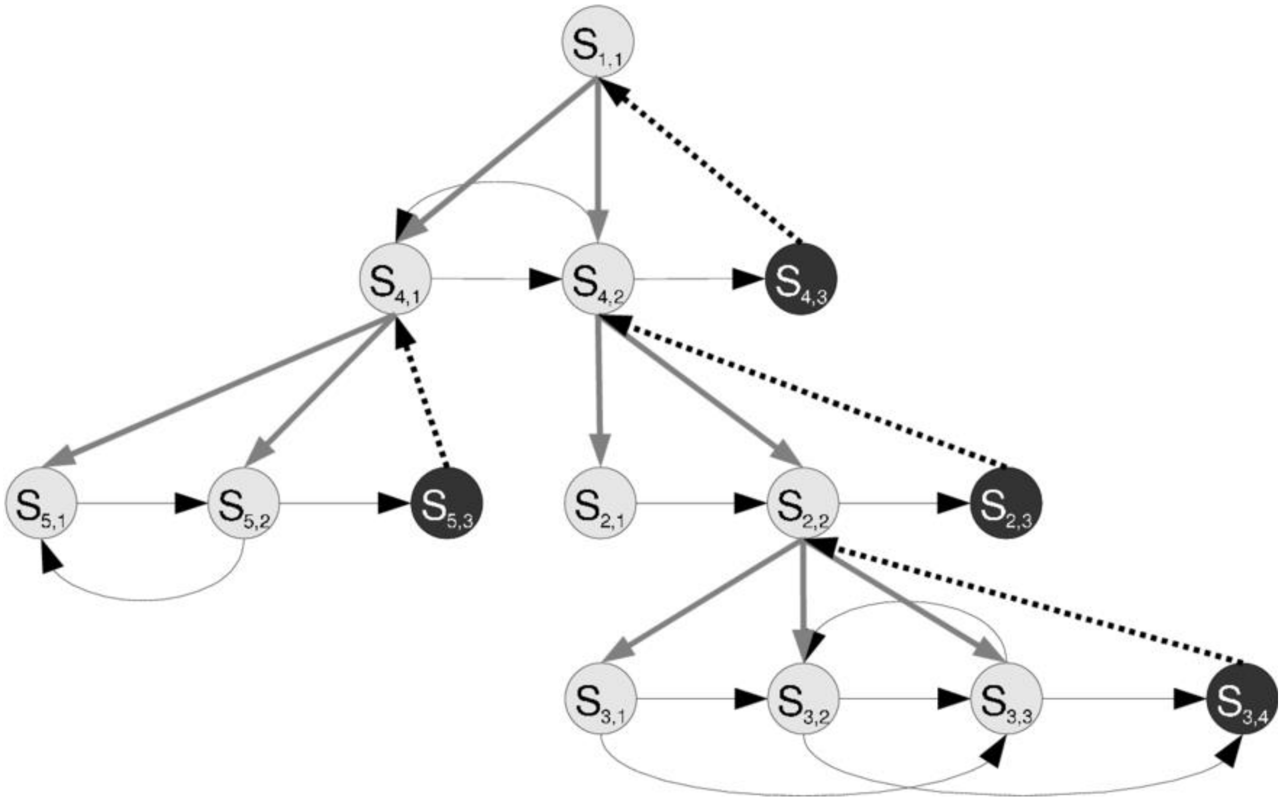
- Ayrık zaman pencelerindeki değerleri ortalamaları ile değiştir
- Bu değerlere sembol ata

- Çok boyut versiyonları mevcut ama …

Şekil Aremu vd.

Örnek: Saklı Markov Modelleri

- Özellikle Hiyerarşik SMMler Öğrenerek hem zamanda boyut küçültme hem de sembol çıkartma yapılabilir

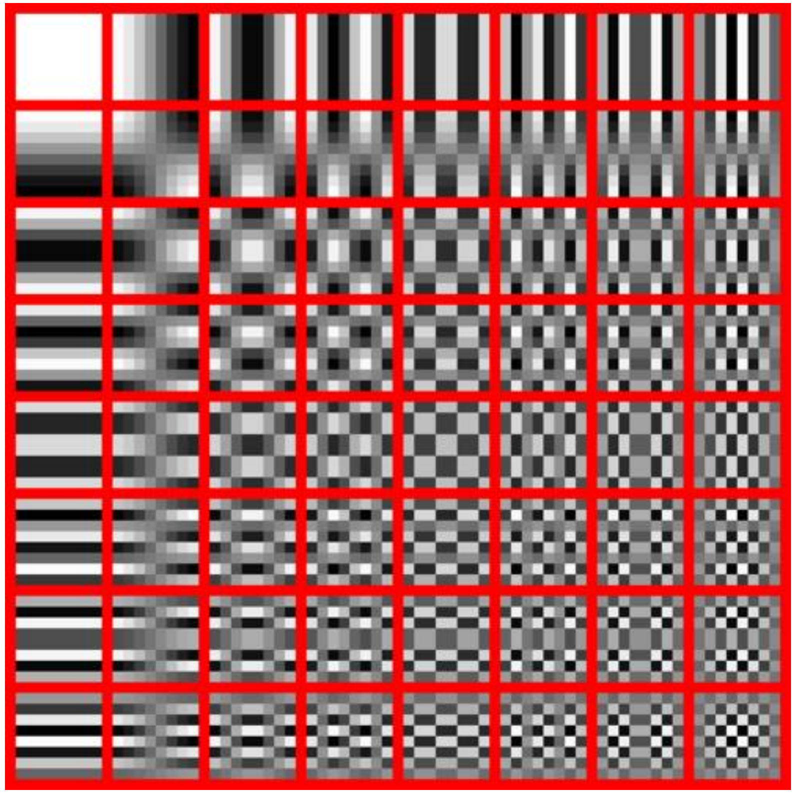


Örnek: Dönüşümler

- Fourier, Dalgacık vb.
- Zaman serisinin dönüşümlerini hesapla
- Katsayılar zaman serisini anlatıyor

- Örnek: Video sıkıştırması!

- Boyut sayısı arttıkça pratikliğini kaybedebiliyor



Wikipedia: JPEG dönüşüm örneği

(Dinamik) Faktör Modelleri

- Hatırlayalım

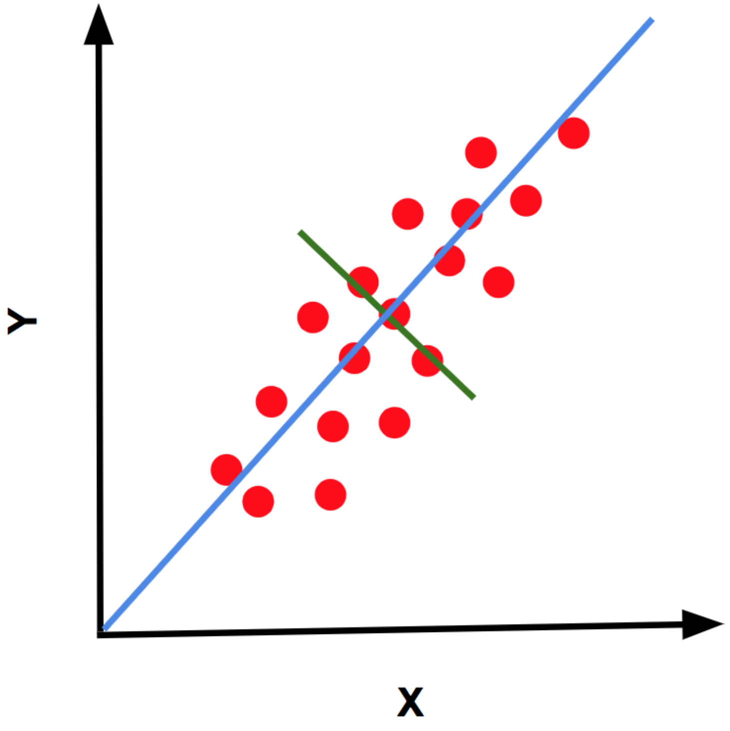
$$X_t = \Lambda F_t + E_t$$
$$F_t = \Phi F_{t-1} + U_t$$

  - F_t boyutu X_t boyutundan daha küçük
  - Daha fazla F gecikmesi de eklenebilir
- Hem Λ hem de F bilinmiyor
  - Λ: Birimdik (ortonormal) varsayılırsa kapsadığı doğrusal uzay özgün
  - Bu sayede kestirim mümkün
- F boyutunun veriden belirlenmesi de mümkün
  - Hatalar ak gürültü olana kadar ara
  - Ya da bir bilgi kriterine göre seç

Doğrudan Boyut Küçültmek

- Gözlemleri özdeş ve bağımsız olarka dağılmış olarak kabul et
  - Not: Zaman serisi analizine ters!
- Bildiğimiz yöntemler ile boyut küçült. Örn:
  - Temel Bileşen Analizi (TBA)
  - UMAP
- Zaman serisi modellemesini küçük boyutta yap

- Dezavantajları:
  - Yüksek boyuttaki zaman bilgisini kaybetme riski
  - Yüksek boyutta kovaryansı kestirimi sorunlu



Fonksiyonel Temel Bileşen Analizi

- Zaman serisi verisinden doğrusal yöntemler *özfonksiyonlar* (eigenfunctions) çıkarmak
- $X(t)$ rassal bir süreç olsun, $\mu(t) = \mathbb{E}[X(t)]$. Otokovaryansı:

$$cov\big(X(s), X(t)\big) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t)$$

  - $\lambda_k > 0$, özdeğerler, ϕ_k: özfonksiyonlar
- Bu süreci temel fonksiyonlar ile gösterebiliriz:

$$X(t) - \mu\big(X(t)\big) = \sum_{k=1}^{\infty} \xi_k \phi_k(t)$$

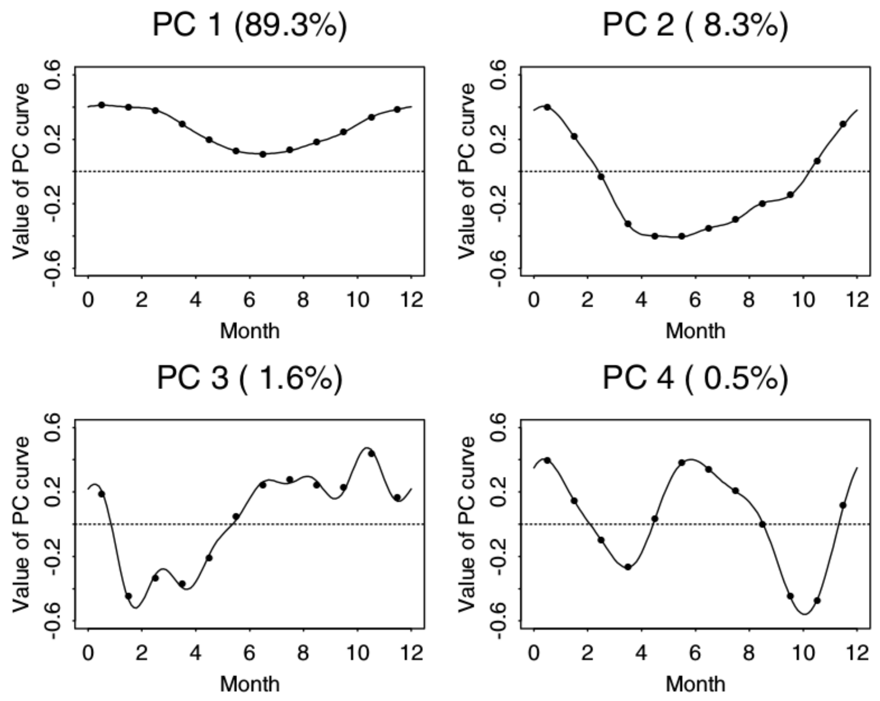
  - ξ_k: temel bileşenler

Fonksiyonel Temel Bileşen Analizi

- Kullanım: m temel bileşen seçtikten sonra:

$$X(t) \approx \mu(X(t)) + \sum_{k=1}^{m} \xi_k \phi_k(t)$$

- Seyrek olmayan serilerde, zamanda düzenlileştirilmiş TBA ile eşdeğer
- Dönüşüm yaklaşımı olarak yorumlanabilir
- Faktör modeli öğrenmek için kullanılabilir



Ramsay ve Silverman

Bir Adım Geri

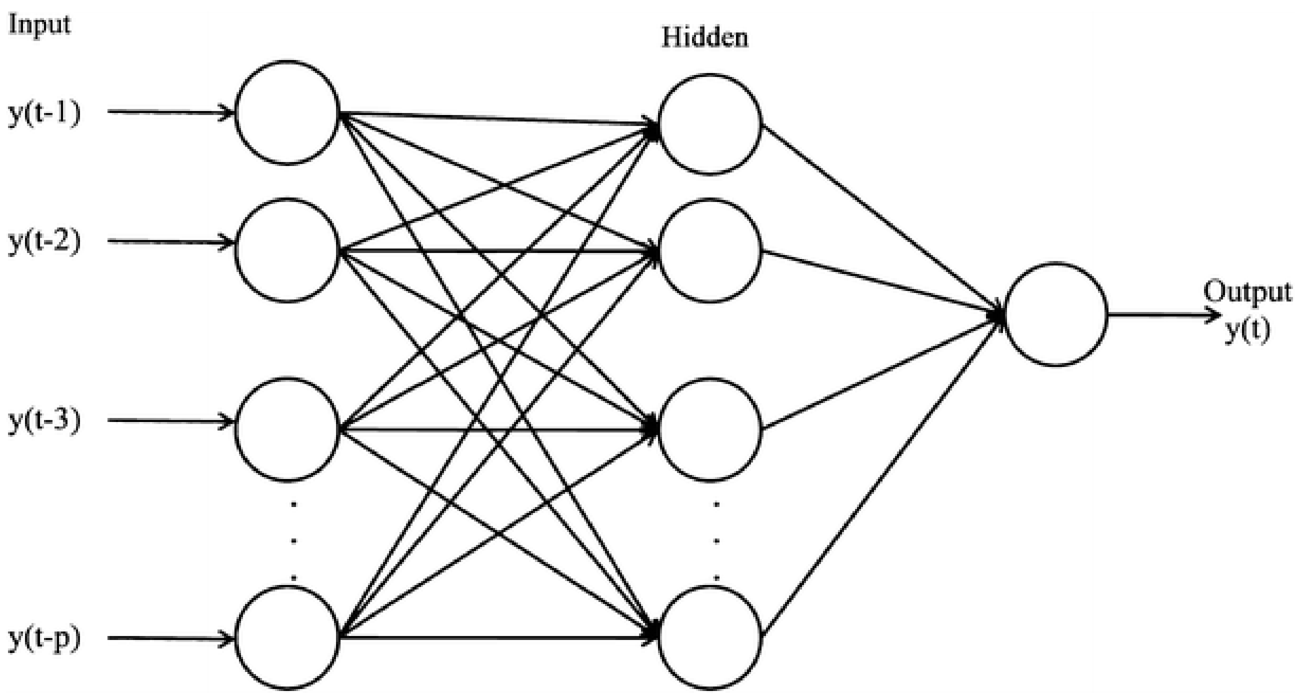
- Daha birçok model var:
  - ARCH, GARCH, TAR, LVARMA …
- Parametresiz modeller de mevcut
  - Gauss Süreçleri
- Herhangi bir gözetimli öğrenme yöntemini, girdi ve çıktılarını ayarlayarak zaman serisi modeline çevirmek de mümkün

$$X_t = f\left(X_{t-1}, X_{t-2} \dots, X_{t-p}, w\right)$$

- Yüksek boyutta gözlemler yapsak bile alttaki olaylar genelde daha düşük boyutlu varsayımı genelde doğru
- Boyut küçültme – gösterim öğrenme – derin modeller
  - Derin modeller doğru mimariler ile çok boyutlu verileri halledebiliyor

Yapay Sinir Ağı Zaman Serisi Mimarileri

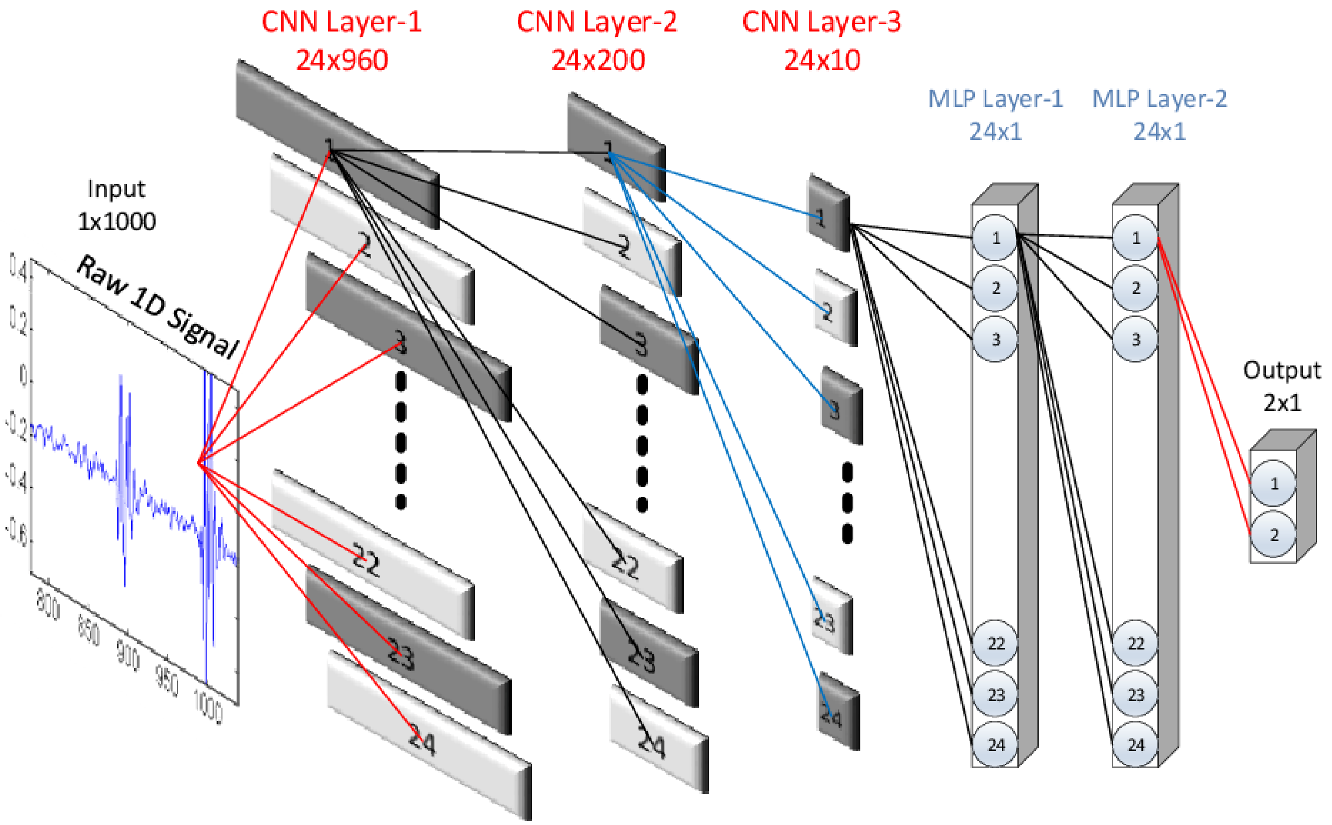
- Zaman Gecikmeli Sinir Ağı:



- Tek filtreli 1 Boyutlu Evrişimsel Sinir Ağı
- Çıktı zaman bağlı başka bir değişken olabilir

Yapay Sinir Ağı Zaman Serisi Mimarileri

- 1 Boyutlu Evrişimsel (Convolutional) Sinir Ağı – Zaman Ekseninde



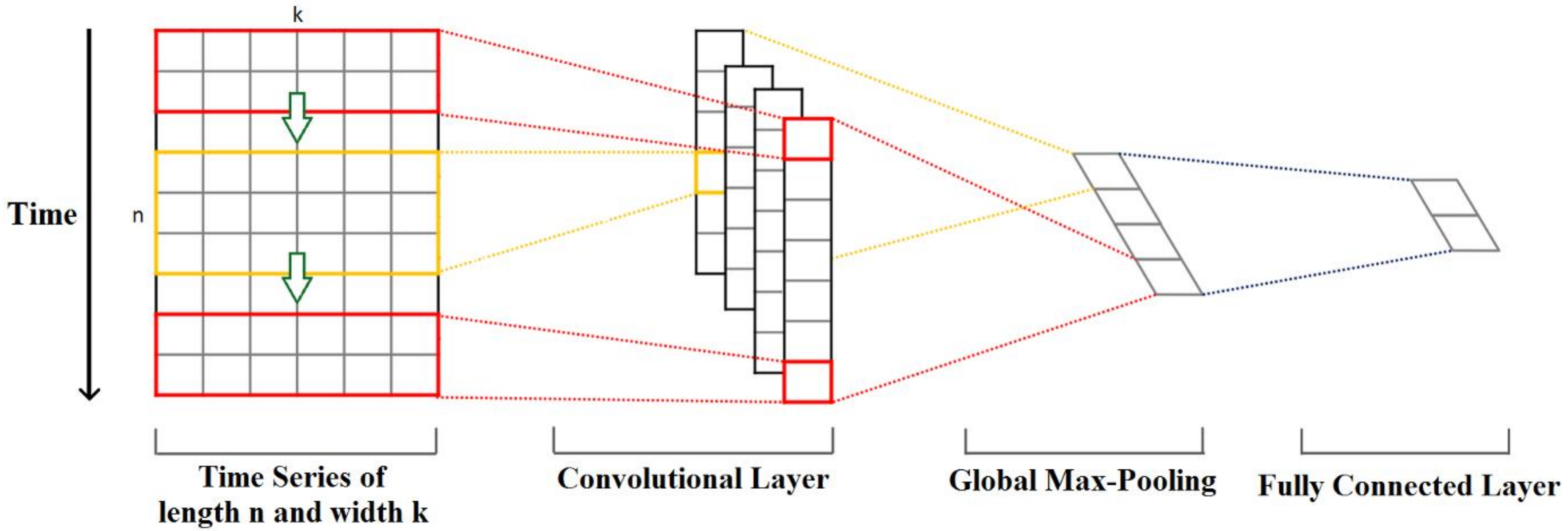
Kiranyaz vd.

- Evrişimsel ağlar, verilen eksende parametre paylaşırlar
- Bu sayede karmaşıklıkları azalır
- Aktivasyon fonksiyonları ve derinlik sayesinde de doğrusal olmayan bağlantıları öğrenirler

- Not: Evirişimsel ağlar, zaman serisi regresyonu için genelde çok iyi performans göstermezler

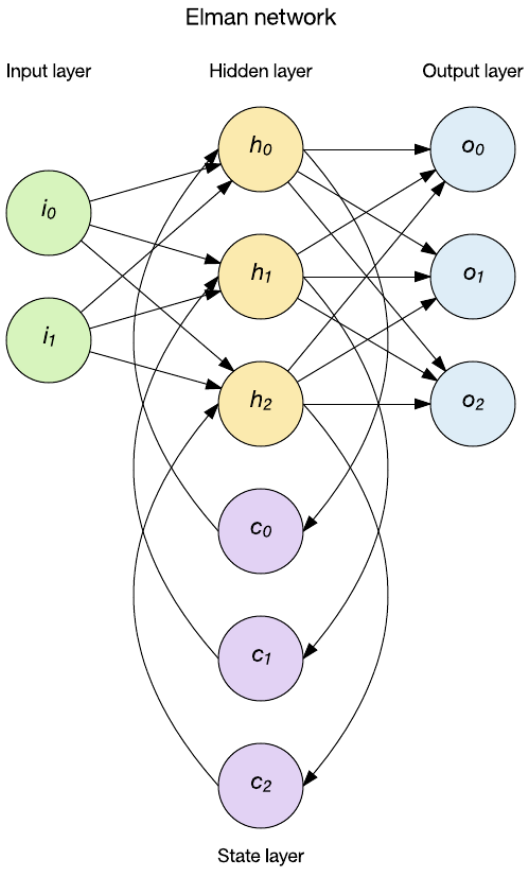
Yapay Sinir Ağı Zaman Serisi Mimarileri

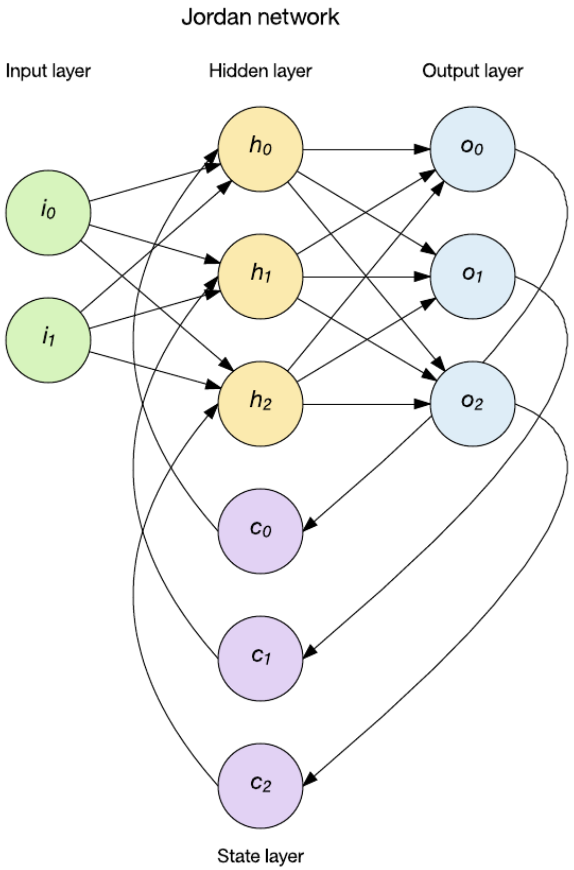
- 1 Boyutlu Evrişimsel Sinir Ağı – Ölçüm Ekseninde



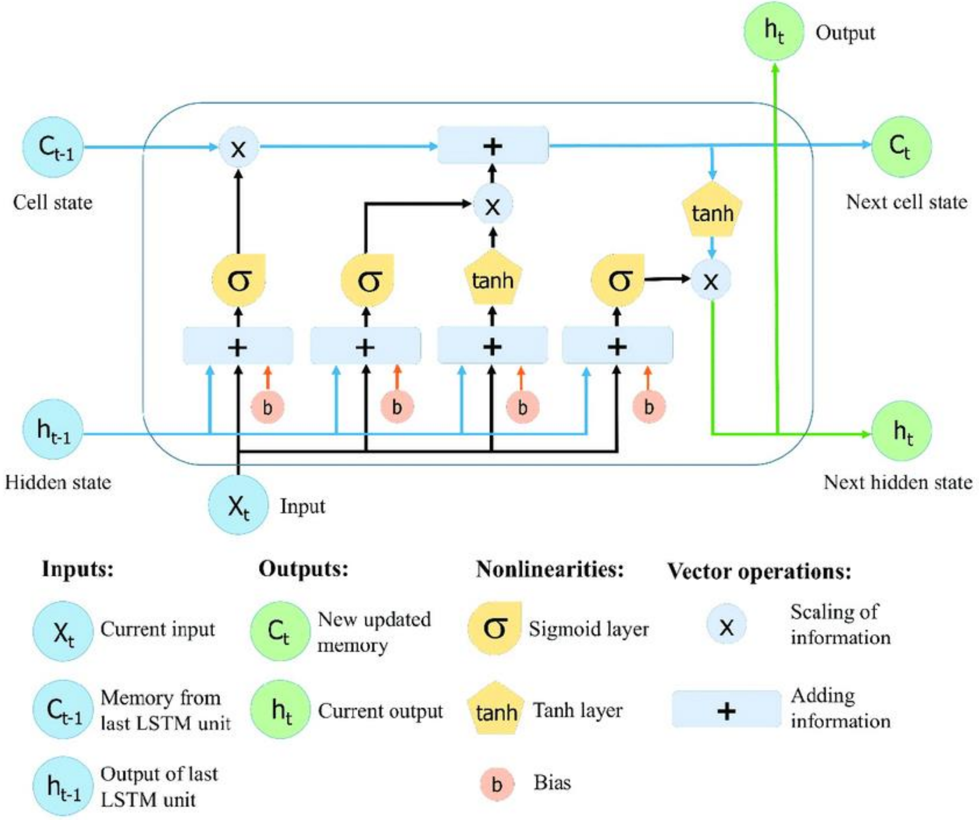
Yapay Sinir Ağı Zaman Serisi Mimarileri

- Özyineli (Recurrent) Ağlar





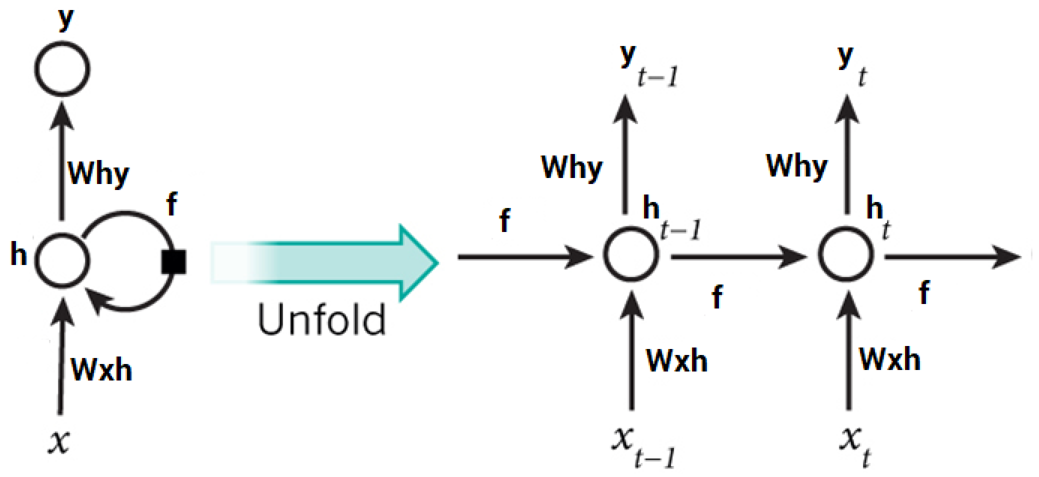
LSTM

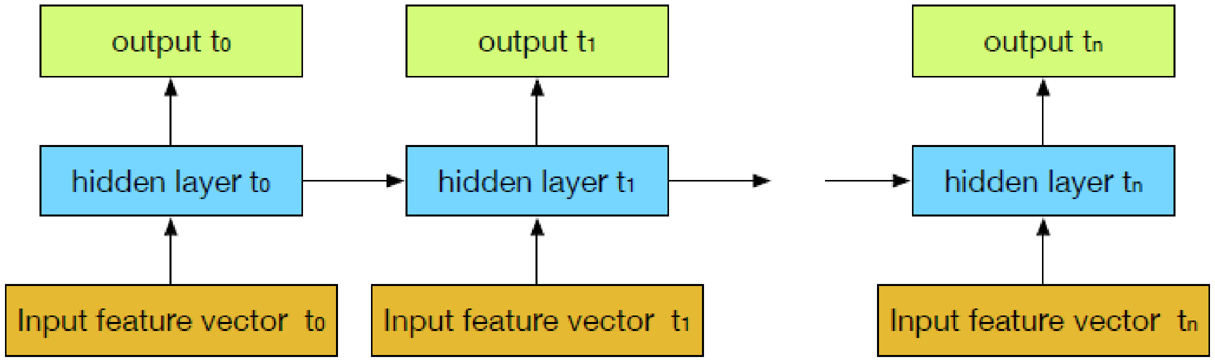


Bir nevi "hafıza" özelliği katıyor

Yapay Sinir Ağı Zaman Serisi Mimarileri

- Özyineli Ağlar





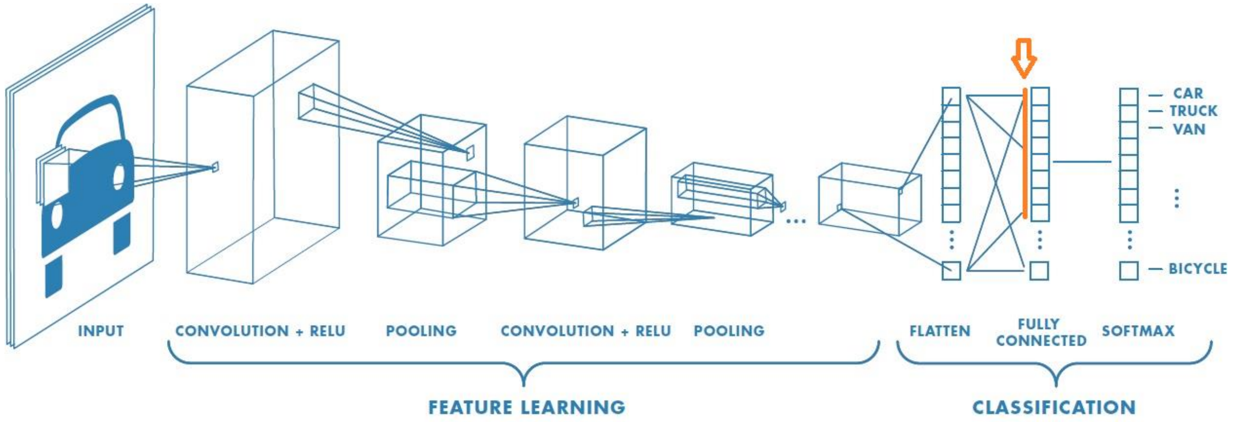
- Özyineli ağlar, zamanda parametre paylaşırlar
- Bu sayede karmaşıklıkları azalır
- Aktivasyon fonksiyonları ve derinlik sayesinde de doğrusal olmayan bağlantıları öğrenirler

- Zaman serisi regresyonunda genelde fena değillerdir.
- Çok boyutlu problemlerde, en genel VAR modellerinden çok daha hızlı sonuç verirler

- Hiyerarşik olarak kullanılabilir
- Özyineli katmanlardan önce evrişimsel katmanlar eklenebilir

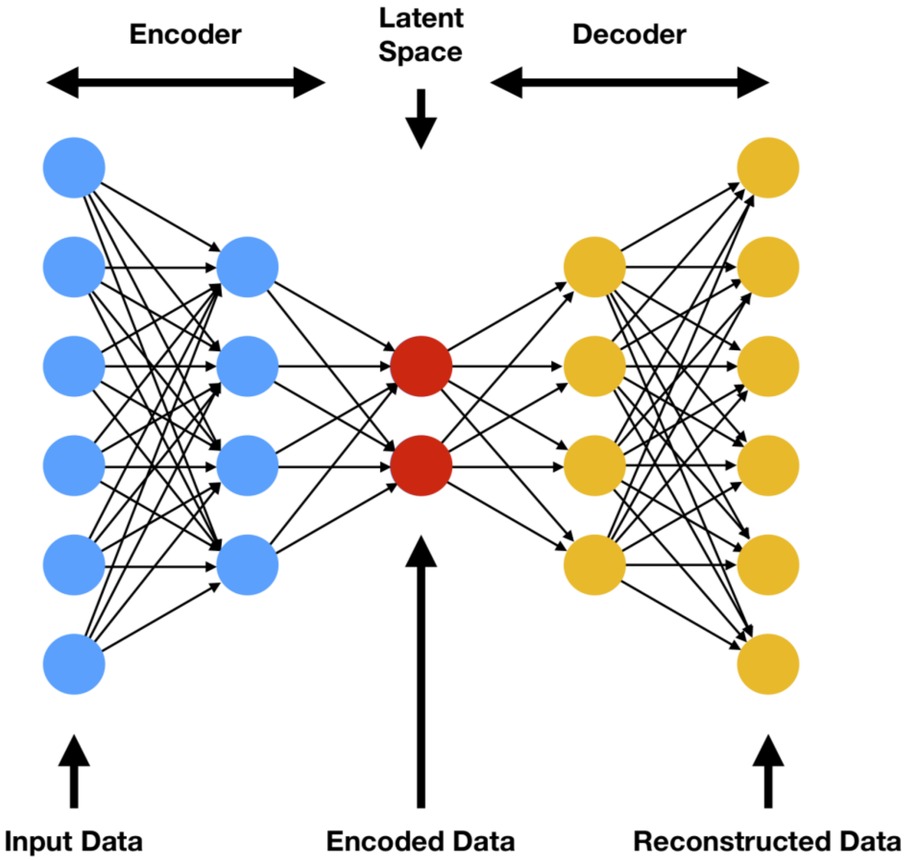
Yapay Sinir Ağları ile Basit Boyut Küçültme

Öğren ve
Tepesini Çıkart



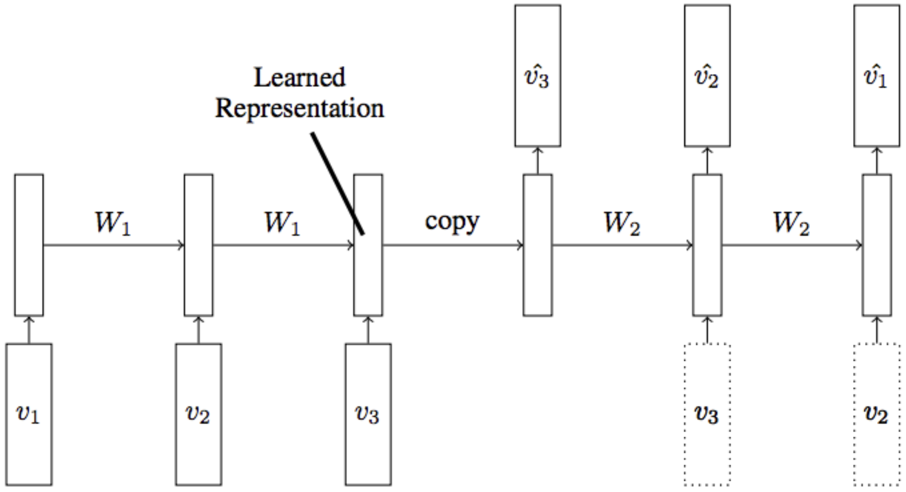
Aslında benzer!

Otokodlayıcı
(Gözetimsiz*)



YS ile Zaman Serilerinde Boyut Küçültme

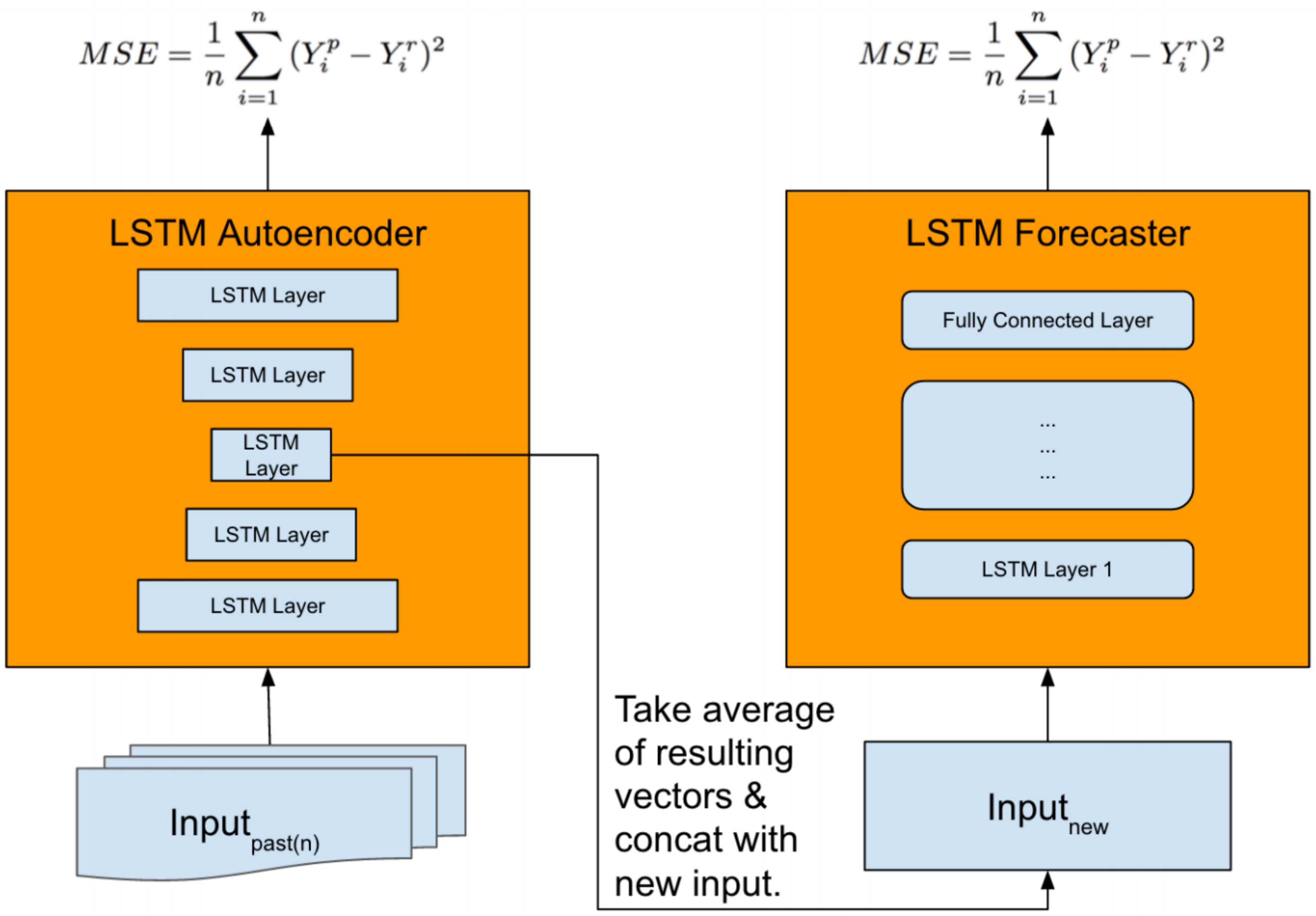
- Basit fikir:
  - Otokodlayıcı kulan
  - Kodlayıcı (ve istenirse çözücü) kısmını zaman serilerine uygun bir model olarak seç
  - Ara katmanı çıktısı düşük boyutlu gösterim olarak al



Srivastava vd.

LSTM – LSTM otokodlayıcı

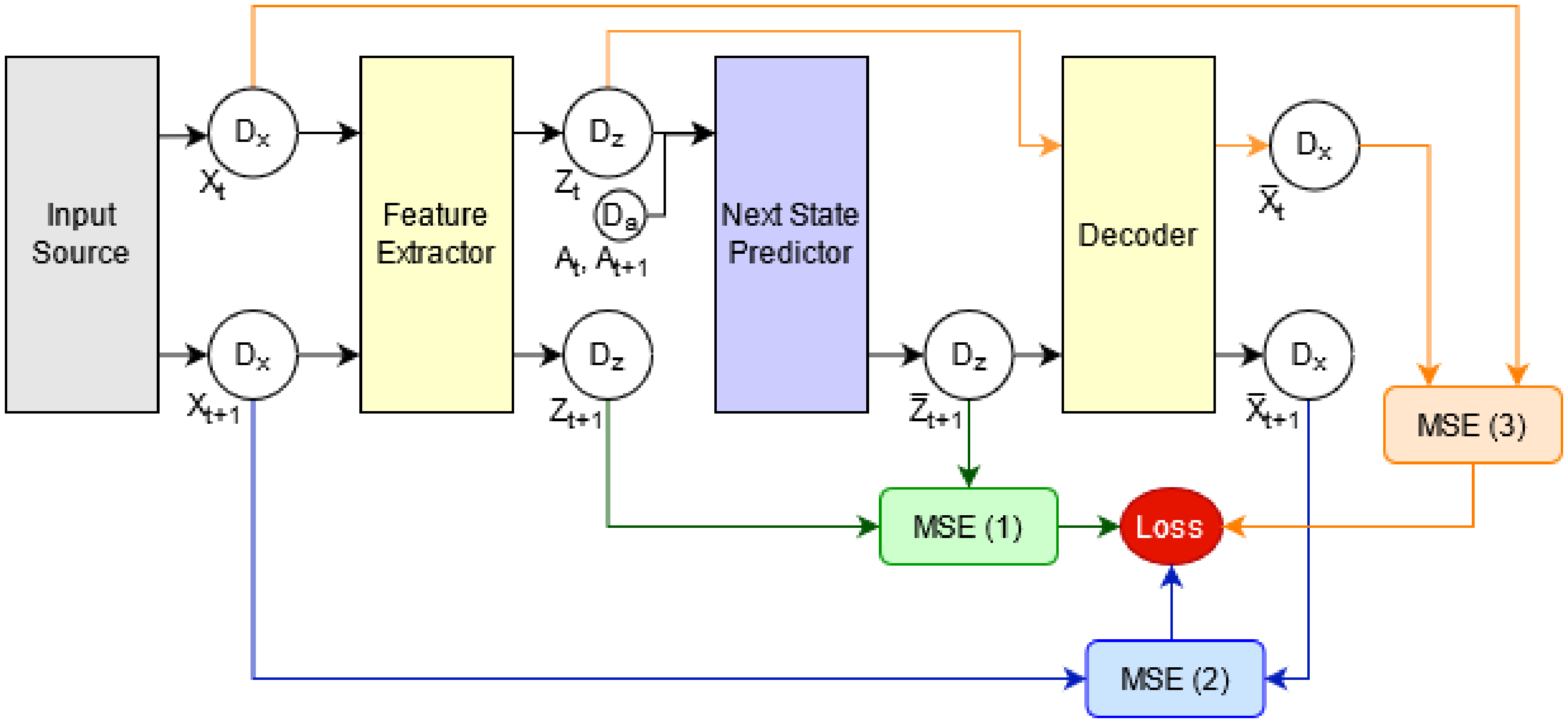
Uber Örneği



Laptev vd.

- Tahmin ve otokodlayıcı iç içe
- Uzun ve çok boyutlu zaman serilerinde ender olay tahmini için klasik yöntemlerden daha iyi çalışmış
- Kısa ve az boyutlu problemlerde klasik yöntemler daha iyi
- Uzun ve az boyutlu problemlerde performans benzer

Düşük Boyutta da Tahmin ve Karar Girdileri

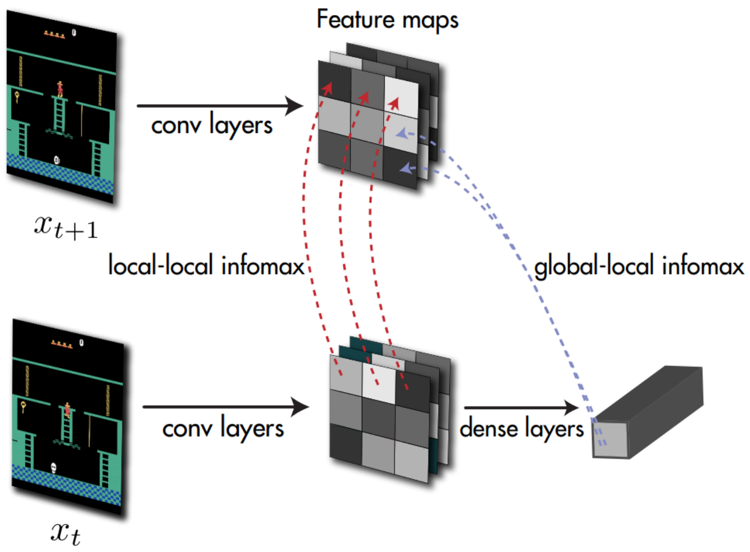


Serteli ve Akgün (deney aşaması devam ediyor)

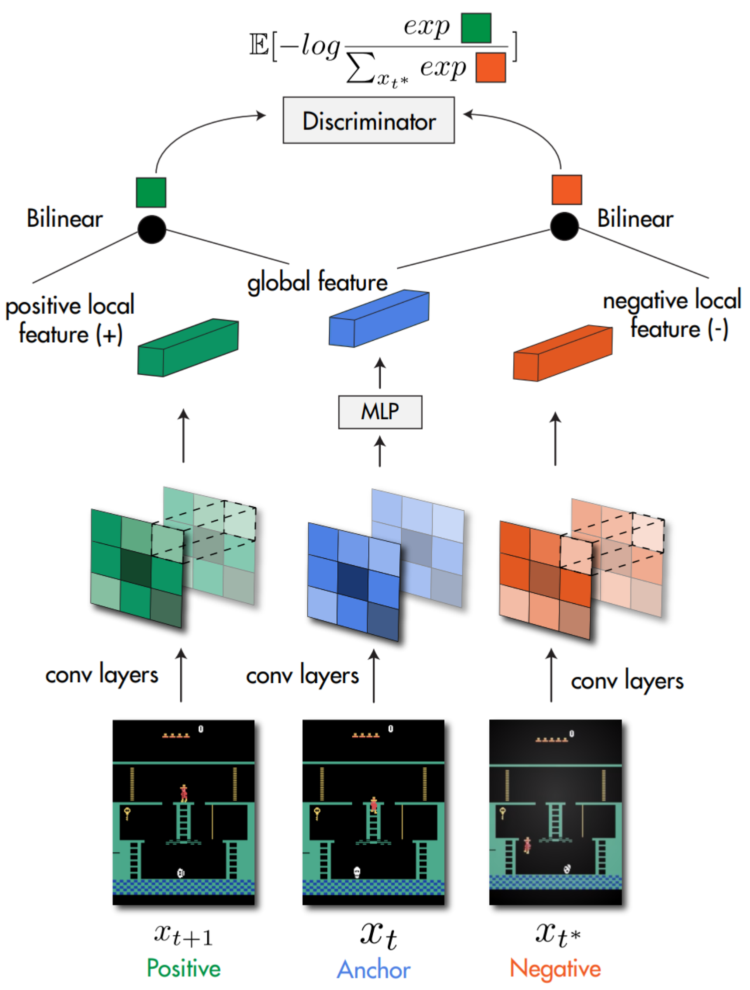
- X: çok boyutlu veri
- Z: düşük boyutlu veri
- A: karar girdisi (örneğin robot aksiyonu)
- Aynı anda:
    - Çok boyut tahmini
    - Düşük boyut tahmini
    - Boyut küçültme
- Hafıza için mavi kutuda özyineli katmanlar kullanılabilir
- "Az" veriden öğrenilebilecek şekilde tasarlandı

Not: Buna benzer mimariler mevcut. Detaylarda bizim uğraştığımız problemlere özelleşmiş kısımları var

Ayrımsal (Contrastive) Yöntemler

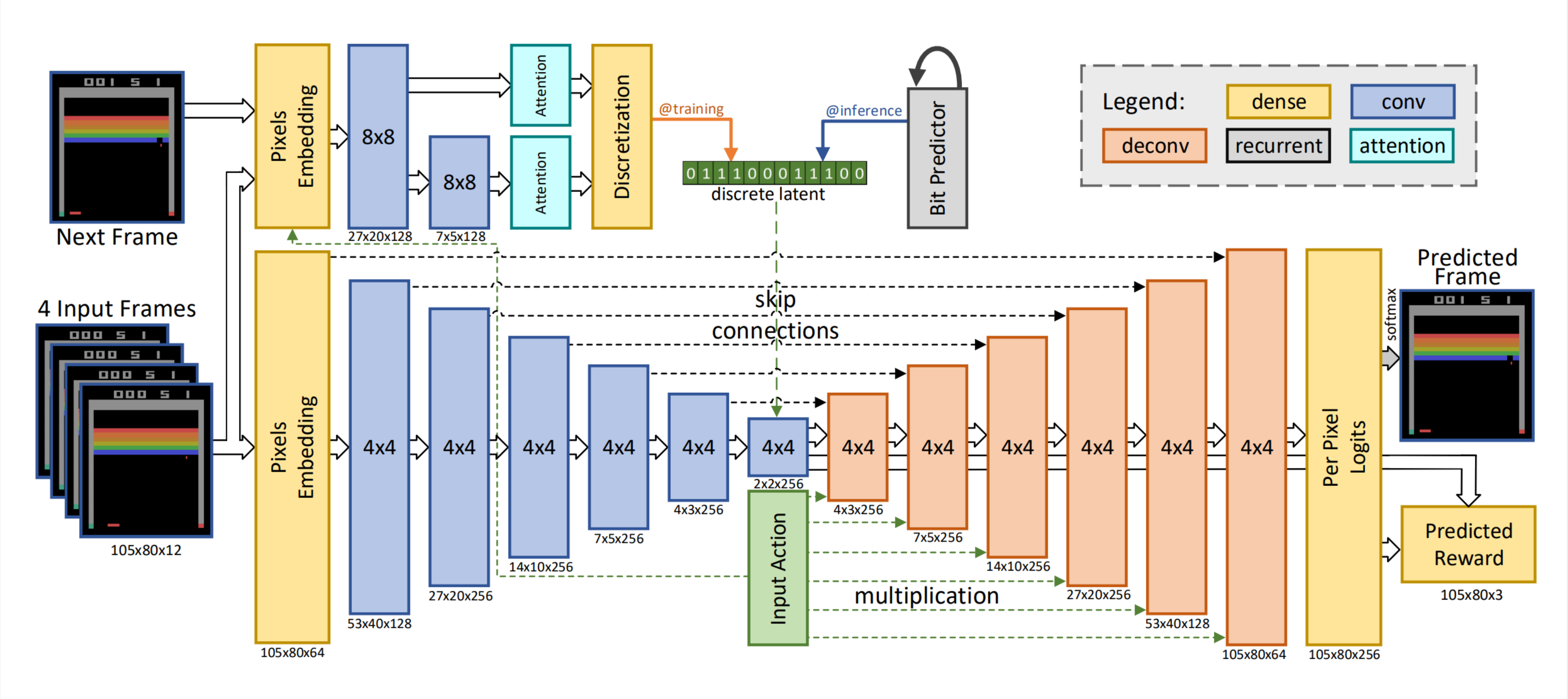


Anand vd.



- Mimari, iki durumu ayırması ve böylece daha iyi gösterim öğrenmesi için zorlanabilir
- Bu örnekte, bir sonraki Atari ekranı, şimdiki Atari ekranından tahmin edilmeye çalışılıyor
- Çok veri gerekli

Pekiştirmeli Öğrenme ile Beraber

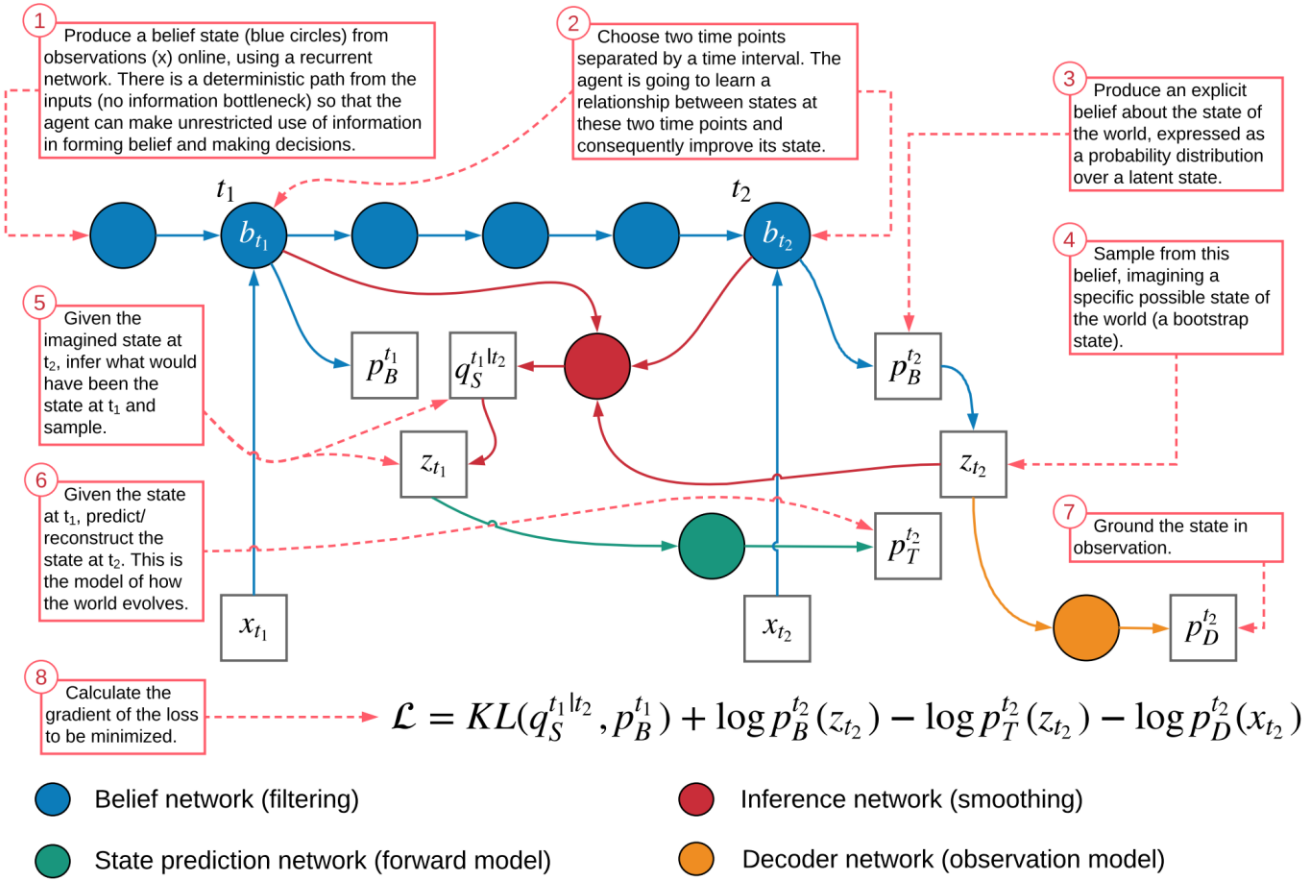


Kaiser vd.

- Küçük boyut aksiyon etkisi için kullanılıyor
- Arka plan çok değişmediği için mantıklı
- Ancak küçük boyut kullanılmıyor

Hem Düşük Boyut Hem Simülasyon

- Zamansal Fark Değişimsel Otokodlayıcı (Temporal Difference Variational Autoencoder - TD-VAE)



- Düşük boyutlu soyut bir gösterim öğren
- Durum ile ilgili belirsizliği temsil et
- Gelecek tahminlerini olasılıksal yap (yani farklı tahminler mümkün)

- Eksik: Aksiyonlar yok

Gregor vd

Neleri Konuşamadık:

- Farklı problem tipleri için yaklaşımlar ve farklı mimariler
  - Saf zaman serileri
  - Kontrol problemleri
  - Dil
  - …
- Derin yaklaşımlarla sembol öğrenme
- Boyut küçültme öncesi ve sonrası uygulama karşılaştırmaları
- Sürekli öğrenmek ve durağan olmayan problemler

Sonuçlar

- Zaman serileri her yerde. Git gide miktar ve boyut olarak da artıyor
- Geleceği tahmin etmek zor, dünya sürekli değişiyor
- Farklı disiplinlerde benzer problemler mevcut
- Derin modeller aşağıdaki koşullarda tercih edilmeli:
    - Yorumlanabilirlikten taviz verebiliyorsak (bunu çok yapıyoruz, gidişat pek iyi değil)
    - Yeterince uzun veri varsa (ya da mimarimiz özelleşmiş ise)
    - Gözlemler çok boyutluysa
    - Aynı veri ile birden fazla problem ile uğraşacaksak
    - Durağan olmayan problemler (sürekli öğrenerek)