



**T.C.
İSTANBUL ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ**



BİTİRME PROJESİ

[DIMENSIONALITY REDUCTION TECHNIQUES]

[ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ]

**ALİ ÖZTAŞ
1357130053**

**DANIŞMAN
[Yrd. Doç. Dr.Mahmut Öztürk]**

[Ocak, 2017]

İSTANBUL

ÖNSÖZ

Bu ödevde Dimensionality Reduction Techniques (Boyutluluk Azaltma Teknikleri) anlatılıp,boyutluluk fazlalığında ne gibi sorunlar olabileceği ve bu sorunların giderilmesinde kullanılan teknikler anlatılacaktır.

Bu çalışmada yardımını esirgemeyen hocam Yrd.Doç. Dr. Mahmut Öztürk'e teşekkür ederim. |

Ocak 2017

|Ali Öztaş|

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	i
İÇİNDEKİLER.....	ii
1. GİRİŞ	1
2.GENEL KISIMLAR.....	2
2.1.EIGENDECOMPOSITION.....	4
2.2 KOVARYANS MATRİSİ.....	5
2.3 APPENDIX A: LINEAR ALGEBRA.....	6
3.1 PRINCIPAL COMPONENT ANALYSIS.....	7
3.2 MULTIDIMENSIONAL SCALING.....	10
3.3 ISOMAP.....	11
3.4 KERNEL PCA.....	13
3.5 SAMMON MAPPING.....	16
4. BULGULAR.....	199
5.TARTIŞMA VE SONUÇ.....	20
KAYNAKLAR.....	21
EKLER.....	23

1. GİRİŞ

Son yıllarda data toplamaktaki ve bunları depolamakta olan gelişmeler birçok bilim dalında bilgilerin fazlalaşmasına önderlik etmiştir. Biyoloji, astronomi, mühendislik, ekonomi, tüketici işlemleri gibi çeşitli alanlarda çalışan araştırmacılar günlük olarak çok büyük gözlemler ve simülasyonlarla karşılaşır. Küçük ve daha gelenekselleşmiş datasetlerinin tersine az önce bahsedilen datasetleri incelenmiştir ve data analizinde yeni sorunlar ortaya çıkarılmıştır. Geleneksel istatistik yöntemler gözlem sayısının fazlalığından bozulsalar da asıl sebep her gözlem ile değişken sayısında da artış görülmektedir. Bir datanın boyutu her gözlemde ölçülen değişken sayısıdır.

Yüksek boyutlu datalar matematiksel olarak da fazla sayıda sorun ve bir kısım fırsatlar ortaya koyar. Yüksek boyutlu datasetlerin sorunlarından bir tanesi, çoğu durumda ölçülen her değişkenin istenen olguda önemli olmamasıdır. Bazı belli hesaplamalı olarak pahalı olan metodlar^[1] yüksek başarı ile yüksek boyutlu data ile ilgili tahmini modeller oluşturabilse de uygulamaların çoğunda orijinal data modellenmeden önce boyutu azaltılır. Örnek olarak aşağıdakiler verilebilir:

- Bir yüz tanıma/sıflama sistemi ele alalım. $[m \times n]$ gri tonlu bir resim satırları birbirine bağlanarak $m \times n$ -boyutlu reel vektörlere dönüştürülebilir. Örnek olarak $m = n = 256$ alınırsa 65536-boyutlu vektörler elde edilir. Sınıflandırma sistemi olarak da Çok Katmanlı Algılayıcı kullanılırsa ağırlıkların sayısı çok fazla olacaktır ve çok büyük bir eğitim verisi gerekecektir. Bundan dolayı boyutu azaltmak gerekir. Kaba olarak resimler küçültülerek daha yönetilebilir bir boyuta getirilebilir daha detaylı çözümler de mevcuttur.

Daha da genel olarak söylenirse, bir veri setinin içsel boyutluluğu asıl olanından daha küçük ise boyutluluk azaltma hesaplama olarak kolaylık sağlamanın yanında datayı daha iyi anlamamızı sağlar. Boyutluluk azaltma özellik çıkarma veya bir kod prosedürü olarak görülebilir ya da genel olarak farklı bir koordinat sisteminde gösterim olarak görülebilir.

2. GENEL KISIMLAR

Boyutluluk azaltma tekniklerini kaba olarak aşağıdaki gibi sınıflandırabiliriz:

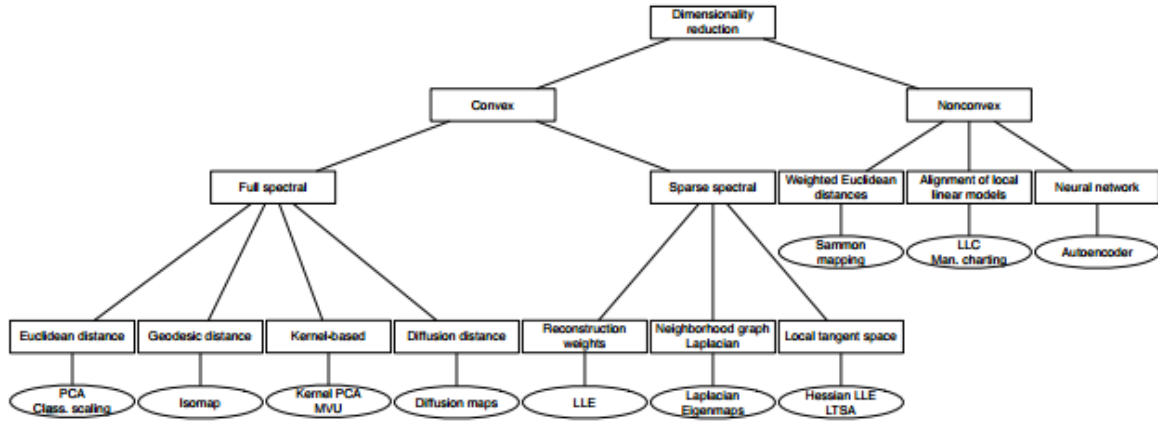
- *Zor boyutluluk azaltma problemleri*, bu problemlerde boyutluluk aralığı çok fazladır. (100-100.000) Daha çok yasaklayıcı küçültmeler kullanılır. Bileşenler genellikle bir büyüklüğün uzayın farklı noktalarında veya farklı zamanlarda ölçülen değerlerinden oluşur. Bu sınıfta içinde resimler de içeren desen tanıma ve sınıflama problemleri bulunur. Konuşma tanıma da bu sınıfın içindedir.
- *Kolay boyutluluk azaltma problemleri*, bu problemlerde data çok yüksek boyutlu değildir ve küçültme çok kısıtlayıcı değildir. Bileşenler genelde farklı değişkenlerin gözlemlenen veya ölçülen değerleridir ve incelenmesi rahattır. Sosyal bilimler, psikoloji gibi çoğu istatistiksel analizler bu sınıfın içindedir.
- *Görüntüleme problemleri*, bu problemlerde data çok yüksek boyuta sahip değildir fakat görüntülemek için 2 veya 3'e indirmek gereklidir. 5 boyuta kadar olan datasetleri göstermek için görüntüleme teknikleri vardır, renkler, yönler, kabartmalar, stereografi veya başka aletler kullanılır fakat basit bir grafiğin çekiciliğine sahip değildirler. Chernoff yüzleri (Chernoff, 1973) daha da fazla boyuta izin verir fakat datayı incelemek ve görüntülemek oldukça zordur.

Zaman değişkeni de işin içine katılırsa iki kategori daha eklenebilir: *static boyutluluk azaltma* ve *zamana bağlı boyutluluk azaltma*. Vektör zaman serilerinde kullanışlıdır. Örnek olarak videolardan kesitler ve devamlı konuşmalar verilebilir.

Boyutluluk azaltma yöntemleri şöyle tanımlanabilir, $n \times D$ bir matris \mathbf{X} ile temsil edilen bir dataset olduğu varsayılırsa, bu matrisin n adet D boyutluluğa sahip data vektörü \mathbf{x}_i vardır. Devam etmek için bu datasetin d içsel boyutluluğunun olduğunu varsayılır. ($d < D$ hatta çoğunlukla $d \ll D$). Burada, matematik terimleriyle içsel boyutluluk, \mathbf{X} datasetindeki noktalar D -boyutlu uzayın içine gömülmüş d boyutluluğuna sahip manifoldun yakınında ya da üstünde bulunmaktadır. Bu manifoldun yapısıyla ilgili hiçbir tahminde bulunulmamaktadır. Boyutluluk azaltma teknikleri, D boyutluluğuna sahip \mathbf{X} datasetini d

boyutluluğa sahip yeni bir dataset olan \mathbf{Y} datasetine dönüştürür. Bu işlem datanın geometrisini koruyarak yapılmaya çalışılır. Genel olarak \mathbf{X} datasetinin, manifoldunun geometrisi veya içsel boyutluluğu d ile ilgili bir bilgi bilinmemektedir. Boyutluluk azaltma problemleri anca belli özelliklerin varsayımı ile çözülebilir (içsel boyutluluk gibi).

Boyutluluk azaltmak için çok sayıda değişken kullanılabilir örnek olarak noktalar arası uzaklıklara bakılabilir fakat 'Geodesic' uzaklığa bakılan teknik ile Öklid uzaklığının kullanıldığı teknik farklıdır. Aşağıdaki şekilde (Şekil-1) bu sınıflandırma verilmiştir. Boyutluluk azaltma, dış-bükey ve dış-bükey olmayan teknikler olarak ikiye bölünmüştür bu ikisi arasındaki ana fark dış-bükey yöntemler datayı bütün olarak incelerken dış-bükey olmayanlar ise lokal olarak inceleyip lokal olarak bir amaç fonksiyonu kullanır.



Şekil-1: Boyutluluk Azaltma Tekniklerinin Taksonomisi

Full Spectral tekniklerde tam matrise 'eigendecomposition' uygulanır ve boyutlar arası kovaryanslar ile birlikte data noktaları arası benzerlikler bulunur. Sparse Spectral tekniklerde ise datanın lokal yapısı korunmaya çalışır ve genelleşmiş bir 'eigenproblem' çözülmeye uğraşılır.

2.1 EIGENDECOMPOSITION

Eigenvectorler ile eigenvalue değerlerini bulmak için birden fazla yol vardır, en yaygın kullanılan yöntem A matrisinin Eigenvector'ü olan u vektörünü aşağıdaki denklem ile tanımlar.

$$Au = \lambda u \quad (1)$$

yeniden yazıldığında aşağıdaki hale gelir:

$$(A - \lambda I)u = 0 \quad (2)$$

λ eigenvector ile alakalı olan eigenvalue idir. Ve aşağıdaki denklem ile bulunur.

$$|A - \lambda I| = 0 \quad (3)$$

Örnek olarak A matrisinin eigenvalue ile eigenvectorlerini bulunsun:

$$A = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$$

3 no'lu denklem uygulanırsa

$$\lambda^2 - 3\lambda - 4$$

Elde edilen denklemden, $\lambda_1 = 4$ ve $\lambda_2 = -1$ değerleri bulunur.

1 no'lu denklem uygulanırsa:

$$\lambda_1 = 4 \text{ ile ilişkili olan eigenvector } u_1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix},$$

$$\lambda_2 = -1 \text{ ile ilişkili olan eigenvector ise } u_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

olarak bulunur.

2.2 KOVARYANS MATRİSİ

Kovaryans matrisi, bir rastgele vektörün elemanlarının ilişkilerinin ifade edildiği matristir. İki eleman arasındaki kovaryans aşağıdaki formül ile bulunur.

Elde:

$X = \begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix}$ matrisi olursa kovaryans matrisi aşağıdaki şekilde bulunur.

$$\sum_{ij} \text{cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] \quad (1)$$

burada

$$\mu_i = E(X_i) \quad (2)$$

Böylece aşağıdaki matrisi elde ederiz.

$$\sum \begin{bmatrix} E[(X_1 - \mu_1)(X_1 - \mu_1)] & \dots & E[(X_1 - \mu_1)(X_n - \mu_n)] \\ \vdots & \ddots & \vdots \\ E[(X_n - \mu_n)(X_1 - \mu_1)] & \dots & E[(X_n - \mu_n)(X_n - \mu_n)] \end{bmatrix}$$

2.3 APPENDIX A : LINEAR ALGEBRA

Bu kısımda bazı yöntemlerde kullanılan lineer cebir teoremleri gösterilecektir.

1. Orthogonal bir matrisin tersi o matrisin transpozuna eşittir
2. A herhangi bir matris olmak üzere " $A^T A$ " ve " $A A^T$ " değerlerinin ikisi de simetriktir.
3. Bir matris yalnızca orthogonal olarak köşegenleştirilebiliyorsa simetriktir.
4. Simetrik bir matris kendisinin orthonormal eigenvectorleri ile köşegenleştirilebilir.

3.1 PCA (Principal Component Analysis)

Principal Component Analysis(PCA) [3,4] boyutluluk azaltmak için kullanılan lineer bir tekniktir yani boyutluluk azaltmayı veriyi daha küçük boyutluluğa sahip lineer bir alt uzaya yerleştirir.Eski olmasına rağmen en çok kullanılan tekniklerden biridir.

PCA datanın az boyutlu gösterimini varyansı olabildiğince maksimum tutarak yapmaya çalışır,bu data için küçük boyutluluğun bir temeli bulunarak yapılır ve bu temelde datanın varyansı maksimize edilmeye çalışır.

Matematik terimleriyle ifade edilirse,PCA maliyet fonksiyonunu ($\text{Trace}(\mathbf{M}^T \text{cov}(\mathbf{X}) \mathbf{M})$) minimumda tutan bir lineer haritalandırma \mathbf{M} 'i bulmaya çalışır ve $\text{cov}(\mathbf{X})$,data matrisi \mathbf{X} 'in kovaryans matrisidir.Bu lineer haritalandırma sıfır ortalamalı datanın d principal eigenvectors (principal components) tarafından oluşturulur.Sonuç olarak PCA aşağıdaki soruyu çözer

$$\text{cov}(\mathbf{X}) \mathbf{M} = \lambda \mathbf{M} \quad (1)$$

Bu eigenproblem d principal eigenvalue λ için çözülür.Data noktaları \mathbf{x}_i 'lerin düşük boyutlu data gösterimleri olan \mathbf{y}_i 'ler, \mathbf{x}_i 'lerin lineer bir temele haritalandırılmasından elde edilir.

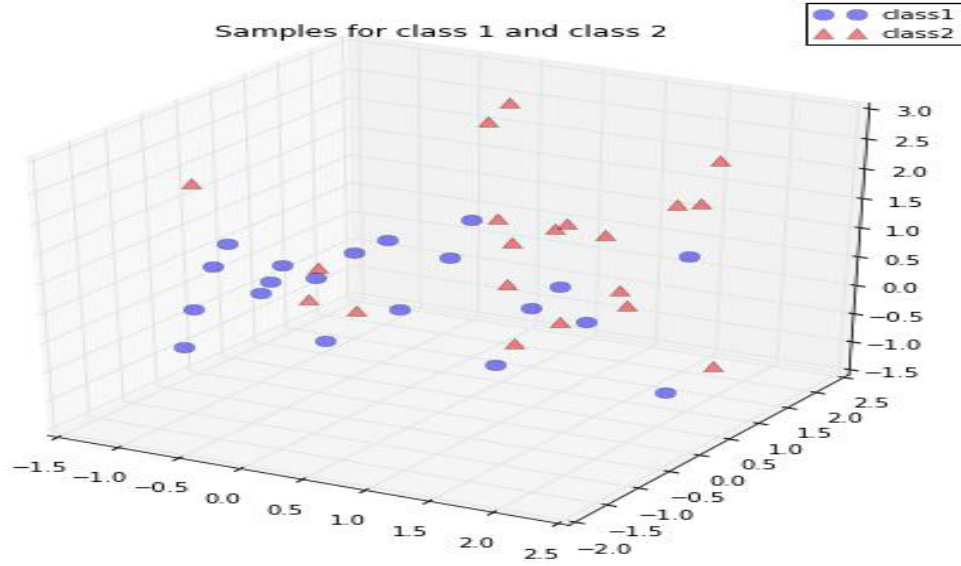
$$\mathbf{Y} = \mathbf{X} \mathbf{M} \quad (2)$$

Bir numaralı ekte verilen sitede PCA'in görsel olarak örneği görülebilir.

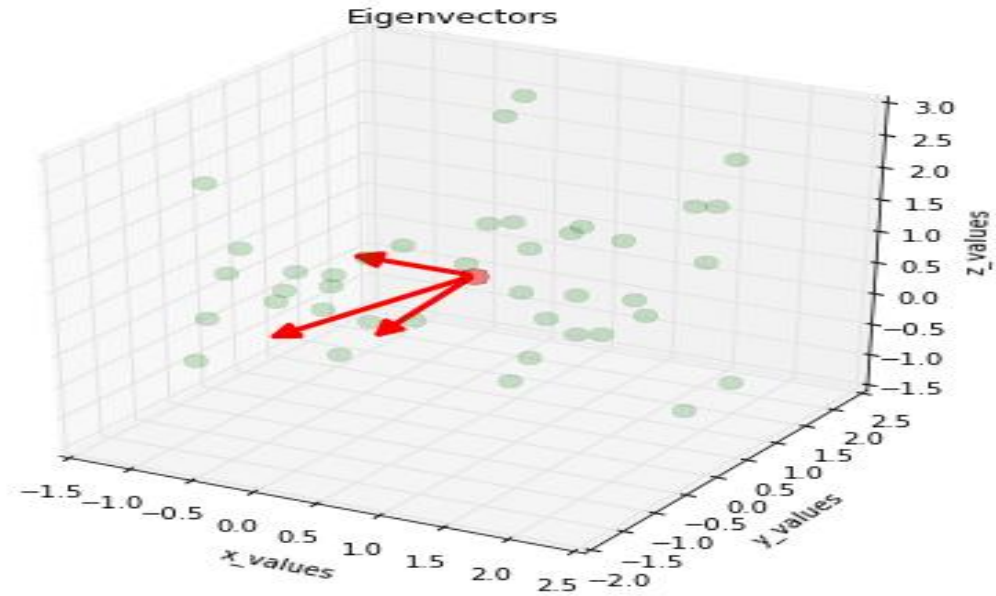
PCA'in zayıflıklarından biri kovaryans matrisinin data noktalarının boyutluluğuyla orantılı olarak büyümesidir ve bununla beraber eigenvectorlerin hesaplanması da oldukça zorlaşır. n sayısının D 'den küçük olduğu zamanlarda 3.2'de tanımlanan MDS yönteminin kullanılmasıyla bu sorun aşılabilir.

3.1.1 PCA ÖRNEK

Bu örnekte 3 boyutlu bir dataset, 2 boyutlu bir datasete çevirilmiştir.Şekil-1'de rastgele dizilip iki sınıfa bölünen noktalar verilmiştir.



Şekil-1



Şekil-2 Eigenvector'lerin ortalanmış halde gösterimi.

Eigenvector 1 : $\begin{bmatrix} -0.84190486 \\ -0.39978877 \\ -0.36244329 \end{bmatrix}$

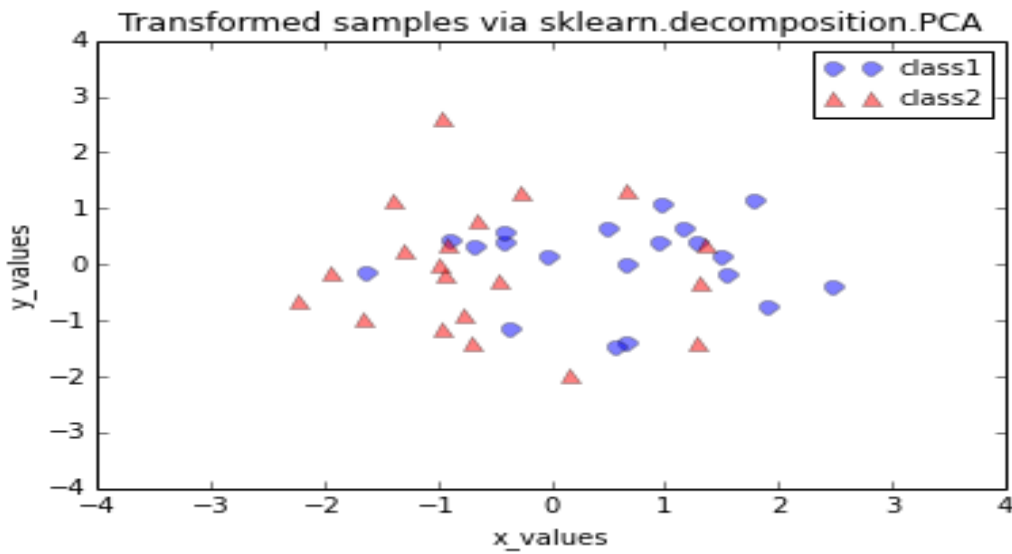
Eigenvector 2 : $\begin{bmatrix} -0.44565232 \\ 0.13637858 \\ 0.88475697 \end{bmatrix}$

Eigenvector 3 : $\begin{bmatrix} 0.30428639 \\ -0.90640489 \\ 0.29298458 \end{bmatrix}$

Bu kısımdan sonra en büyük eigenvalue'ya cevap veren eigenvectorler seçilir. \mathbf{W} matrisi ile adlandırılır ve son olarak $\mathbf{W}^T * \mathbf{x}$ yapılarak yeni gösterim aşağıdaki şekilde(Şekil-3) elde edilir.

$\mathbf{W} =$

$$\mathbf{W} = \begin{bmatrix} -0.84190486 & 0.30428639 \\ -0.39978877 & -0.90640489 \\ -0.36244329 & 0.29298458 \end{bmatrix}$$



Şekil-3

3.2 METRIC MULTIDIMENSIONAL SCALING

Boyutluluk azaltmaya alternatif bir perspektif olarak MDS gösterilebilir. MDS'de orijinal yüksek boyutlu uzayı, düşük boyutlu uzayda gösteren klasik bir yaklaşımdır ama bunu ikili uzaklıkları olarak yapmaktadır. PCA ile matematik olarak farklı olsalar da sonuç olarak benzerlikleri vardır ve lineer bir yerleştirme söz konusudur.

Bir $[t \times t]$ boyutunda D matrisi (uzaklık matrisi) simetrik ise, $d_{ii} = 0$ ve $d_{ij} > 0$.

Uzaklık matrisi D bulunduğundan sonra MDS t sayıda data noktasını (y_1, \dots, y_t) bulmaya çalışır. Ve Metric MDS'in aşağıdaki denklemini minimize ettiği kabul edilir

$$\sum_{i=1}^t \sum_{j=1}^t (d_{ij}^{(X)} - d_{ij}^{(Y)})^2 \quad (3)$$

Yukarıdaki eşitlikte $d_{ij}^{(X)} = \|x_i - x_j\|^2$ ve $d_{ij}^{(Y)} = \|y_i - y_j\|^2$. Uzaklık matrisi $D^{(X)}$ matrisi, $X^T X$ in iç çarpımlarının kerneline dönüştürülebilir.

$$X^T X = \frac{1}{2} H D^{(X)} H \quad (4)$$

$H = I - \frac{1}{t} e e^T$ ve ' e ' tüm 1 lerin sütun vektörüdür. Bu eşitlikler sayesinde (3) numaralı denklem aşağıdaki şekilde indirgenebilir

$$\sum_{i=1}^t \sum_{j=1}^t (x_i^T x_j - y_i^T y_j)^2 \quad (5)$$

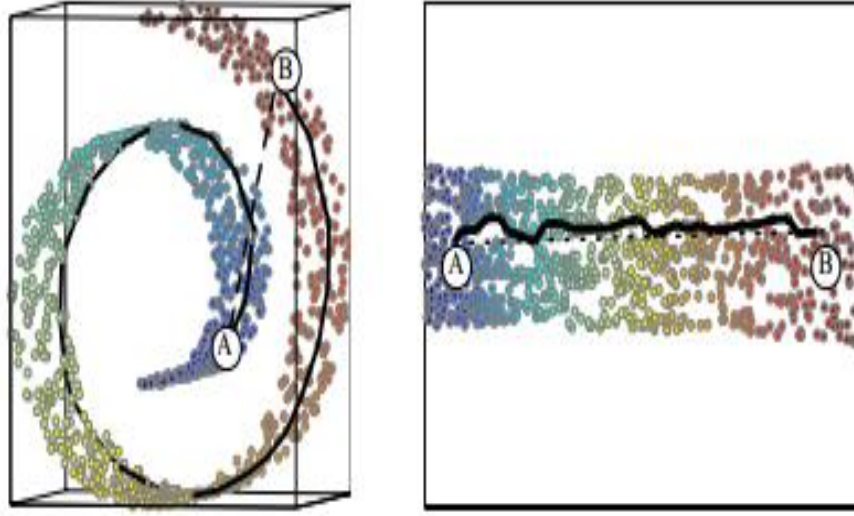
Son olarak çözüm $Y = \Lambda^{1/2} V^T$ olarak gösterilebilir [3]. Burada V , $X^T X$ 'in en büyük d eigenvalue'suna cevap veren eigenvectorler ve Λ ise $X^T X$ 'in en büyük d eigenvalue değerleridir.

PCA ve MDS birçok alanda başarılı olmuştur örnek olarak yüz tanıma [6], para sınıflandırma [7] ve sismik seri analizleri [8] verilebilir.

3.3 ISOMAP

MDS algoritması birçok alanda başarılı olmuştur fakat çiftler arası Öklid uzaklığını kullanması ve data noktalarının komşuluk dağılımını dikkate almamasından dolayı zarar görmektedir. Eğer yüksek boyutlu data eğrilmiş bir manifoldun üstünde ya da yakınında yerleştirilmişse, Swiss roll dataseti gibi [9] MDS, Öklid uzaklığını aldığı için aslında uzak olan iki noktayı yakın olarak kabul edebilir. Isomap [9] tekniği bu sorunu Öklid uzaklıkları yerine Geodesic uzaklığı koruyarak çözen bir algoritmadır. İki nokta arası Geodesic uzaklık manifoldun üzerinde ölçülür. Isomap algoritmasında data noktaları \mathbf{x}_i 'ler arası uzaklık bir komşuluk grafiği G oluşturularak oluşturulabilir, bu grafikte her \mathbf{x}_i data noktası kendisine en yakın k komşusu \mathbf{x}_{ij} ($j = 1, 2, \dots, k$) ile bağlıdır. Bu grafikte iki nokta arası en kısa yol geodesic uzaklığı verir ve Dijkstra'nın veya Floyd'un en kısa yol algoritmaları [10, 11] ile kolaylıkla bulunabilir. \mathbf{X} 'in içindeki data noktaları arasındaki geodesic uzaklıklar bulunduktan sonra ilişkili geodesic uzaklık matrisine yerleştirilir. Büyük boyutlu \mathbf{x}_i 'lerin küçük boyutlu gösterimleri olan \mathbf{y}_i 'ler, geodesic uzaklık matrisine MDS (bkz. 3.2) uygulanmasıyla bulunur.

Isomap algoritmasının önemli zayıflıklarından biri topolojik kararsızlıktır. [12] Isomap komşuluk eğrisi G 'de hatalı bağlantılar oluşturabilir. Zayıflıklardan bir tanesi de Isomap algoritması, manifold konveks değilse başarısız olabilir. Bu zayıflıklara rağmen biyomedikal data gösterimi [13], kafa yönü tahmini [14] ve odun teftişi [15] alanlarında başarıyla kullanılmaktadır.



Şekil 5

Solda: Swiss roll data setinden rastgele seçilen A ve B noktaları arasındaki Geodesic uzaklıklarının karşılaştırılması verilmiştir. Öklid uzaklığı A ve B arası düz bir çizgi ile hesaplanırken Geodesic uzaklık sadece $k=12$ kadar komşu ile bağlantıyı devam ettiren en kısa yolun bulunmasıyla hesaplanmıştır.

Sağda: Swiss roll data setinin $n = 1024$ sayıda giriş ile Isomap tarafından hesaplanan düşük boyutlu gösterimi vardır. Çıkıştaki Öklid uzaklıkları girişteki Geodesic uzaklığa karşılık gelmektedir.

3.4 KERNEL PCA

Kernel PCA(KPCA),klasik PCA'in kernel fonksiyonlarının[16] kullanılarak yeniden formulize edilmesidir.Son yıllarda lineer tekniklerin 'kernel' ile yeniden formulize edilmesi Destek Vektör Makineleri[17] gibi başarılı tekniklerin oluşmasını sağlamıştır.KPCA,kovaryans matrisinin birincil eigenvectorleri yerine kernel matrisinin birincil eigenvectorlerini hesaplar.Kernel uzayında PCA uygulanması,KPCA'e lineer olmayan haritalandırmaları yapma özelliğini getirir.

Kernel PCA \mathbf{x}_i noktalarının kernel matrisi K 'yı hesaplar.Kernel matrisinin girişleri aşağıdaki gibi tanımlanır.

$$k_{ij} = \kappa(x_i, x_j) \quad (6)$$

κ burada kernel fonksiyonudur[17],pozitif-yarı tanımlı herhangi bir kenel \mathbf{K} veren herhangi bir fonksiyon olabilir.Kernel matrisi K aşağıdaki formül ile ortalınır.

$$k_{ij} = \frac{1}{2} \left(k_{ij} - \frac{1}{n} \sum_l k_{il} - \frac{1}{n} \sum_l k_{jl} + \frac{1}{n^2} \sum_{lm} k_{lm} \right) \quad (7)$$

Bu merkezleme işlemi PCA'deki özelliklerin ortalamasını noktalardan çıkarmaya denk gelir.Sonuç olarak kernel fonksiyonu ile tanımlanan özellik uzayındaki data sıfır ortalamalıdır.Merkezlenmiş kernel matrisinin(K) birincil d eigenvectorleri olan \mathbf{v}_i ler bulunur.Kernel matrisi tarafından oluşturulan özellik uzayının kovaryans matrislerinin eigenvectorleri (a_i) artık hesaplanabilir çünkü kernel matrisinin eigenvectorleri ile bağlantılıdır.Aşağıdaki denklem ile hesaplanır.

$$\mathbf{a}_i = \frac{1}{\sqrt{\lambda_i}} \mathbf{v}_i \quad (8)$$

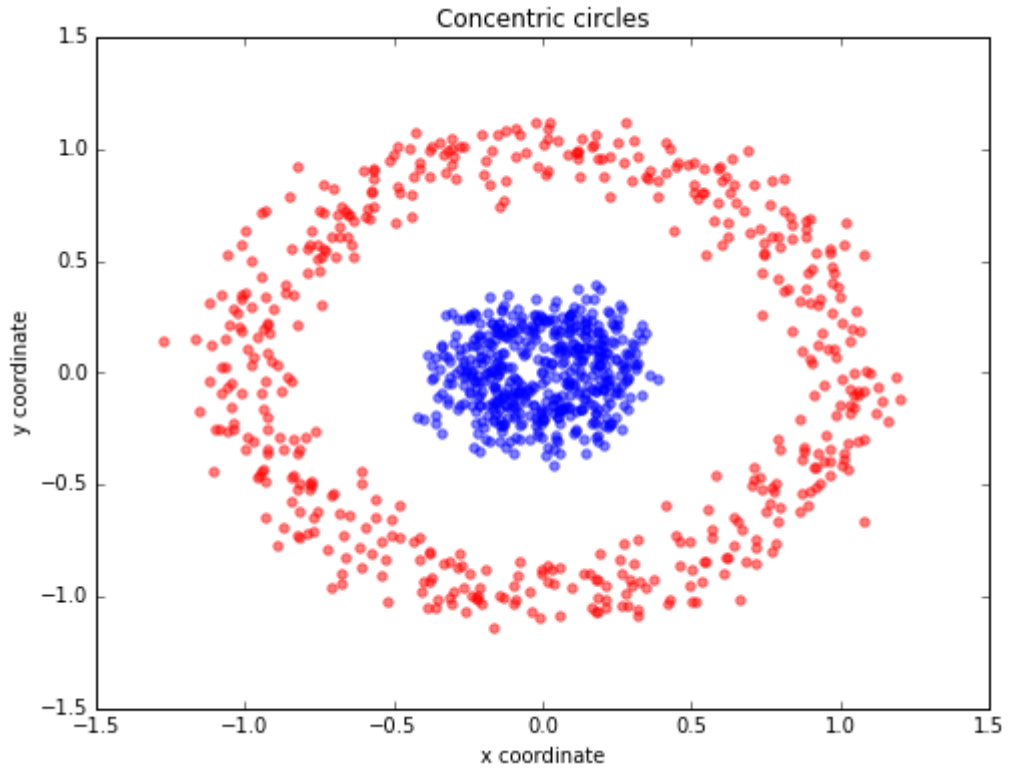
Düşük boyutlu data gösterimini elde etmek için datanın, kovaryans matrisinin eigenvectorleri olan a_i 'lerin üstüne izdüşümü alınır.İzdüşümün sonucu (düşük boyutlu data gösterimi \mathbf{Y}) aşağıdaki formül ile verilmiştir.

$$y_i = \left\{ \sum_{j=1}^n a_1^{(j)} \kappa(x_j, x_i), \dots, \sum_{j=1}^n a_d^{(j)} \kappa(x_j, x_i) \right\} \quad (9)$$

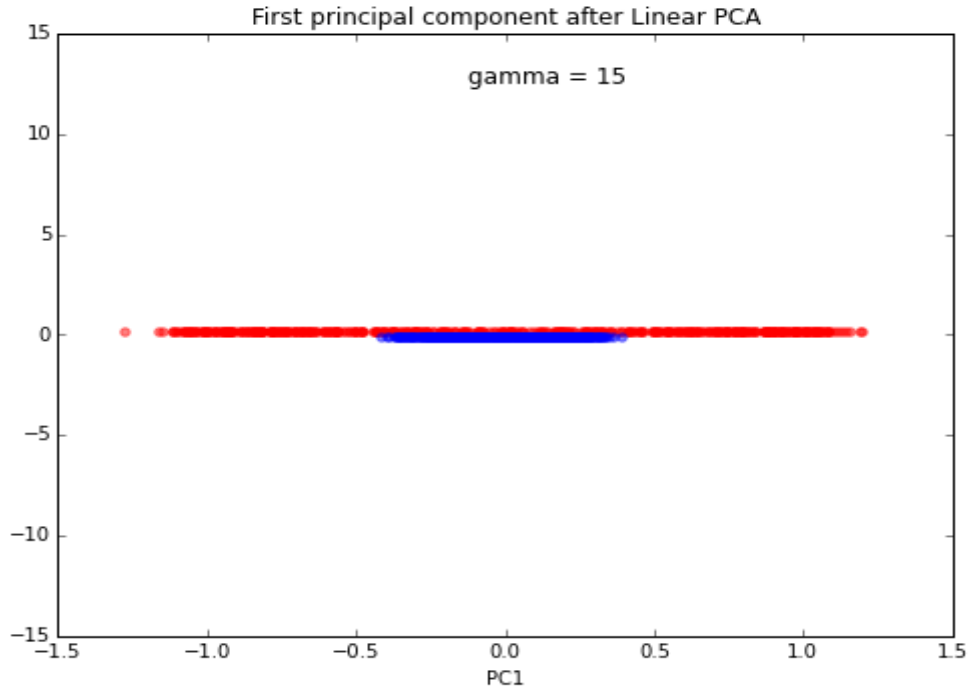
(9) numaralı eşitlikteki $a_1^{(j)}$, a_1 vektörünün j . değerine karşılık gelmektedir ve κ ise kernel matrisinin hesaplanmasında da kullanılan kernel fonksiyonudur.

KPCA'nin önemli bir zayıflığı kernel matrisinin datasetinde bulunan içeriklerin karesiyle orantılı olmasıdır. Ve büyük ilişkili uzaklıkları alması da MDS ile ortak bir zayıflığıdır. Bunlara rağmen yüz tanıma[18], ses tanıma[19] ve tuhaflik bulma[20] gibi alanlarda kullanılmıştır.

Örnek olarak python programlama dilinin ek paketlerinden olan scikit-learn' tarafından sağlanan bir özellik olan ortak merkezli rastgele gürültü ile oluşturulan dairede klasik PCA ve Gauss kerneli kullanılarak yapılan KPCA karşılaştırılırsa:

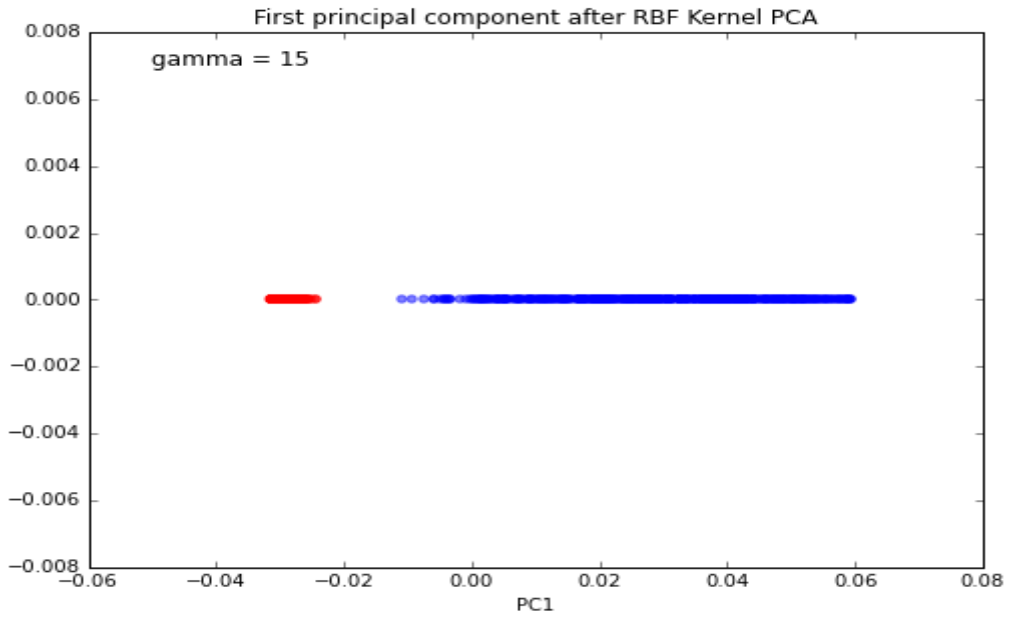


Şekil 6: Rastgele oluşturulan eş merkezli daireler



Şekil 7: PCA uygulandıktan sonra bulunan ilk principal component için düzeltilmiş alt uzay

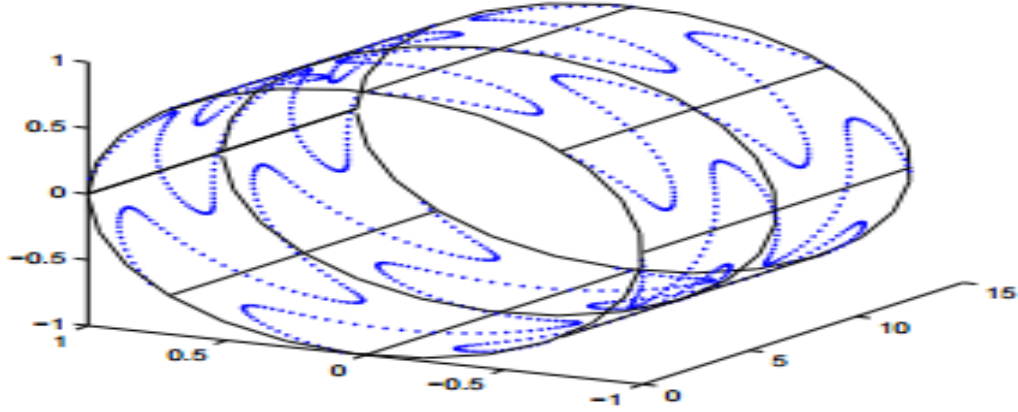
PCA ile üretilen sonuç iki sınıfın lineer olarak iyi ayrıldığı bir alt uzay sunamaz iken Gauss kerneli ile hesaplanan KPCA'nın principal component'i gözlemlendiğinde çok daha iyi bir ayrım yapıldığı görülür.



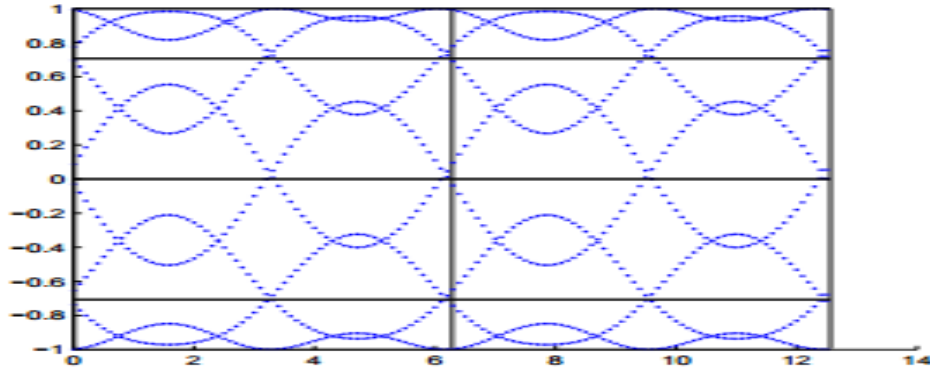
Şekil 8: Kernel PCA uygulandıktan sonra yapılan gösterim

3.5 SAMMON MAPPING

Boyutluluk azaltmak için en basit teknik lineer iz düşümdür örnek olarak PCA'de birincil bileşenler(principal components) üzerine iz düşüm yapılır.PCA orijinal varyansı dönüştürülmüş data setinde maksimum tutmaya çalışsa da karmaşık yapıları koruyamaz[21].Örnek olarak Şekil 9 ve Şekil 10 verilebilir.



Şekil 9: Sekiz tane sinüzoidin bir silindir yüzeyinde gösterimi olan örnek üç boyutlu bir dataset.--Data iki boyutlu bir manifold üstünde olduğu için iki boyutlu olan ve yapısal detayları koruyan bir yapı elde edebilmeliyiz



Şekil 10: PCA uygulamasının sonucu.--Sinüzoidler hala kesişse de topoloji korunmamıştır.

Sammon Mapping algoritması bu yapıları korumak için kullanılan lineer olmayan bir tekniktir.Bu algoritma bir haritalandırma veya dönüşüm fonksiyonu ile işlem yapmaz.Bunun yerine dönüşümün ne kadar iyi yapıldığını ölçer.Başka bir deyişle, orijinal dataya bir

haritalandırma uygulanmamaktadır bunun yerine ilk data setin yapısına olabildiğince yakın olan yeni bir dataseti oluşturulur.

Orijinal dataset L boyutlu bir uzayda N sayıda vektör ile X_i , ($i=1,...,N$) olarak tanımlanır.İstenilen şey bu seti d boyutlu uzayda gösterip Y_i vektörlerini vermektir.Basit olması için d_{ij} 'yi Y_i ve Y_j arası uzaklık ve benzer olarak d_{ij}^* 'yi ise X_i ve X_j arası uzaklık olarak gösterilir.(Sammon bu uzaklıkları ölçek olarak Öklid olarak alsa da bu bir zorunluluk değildir).

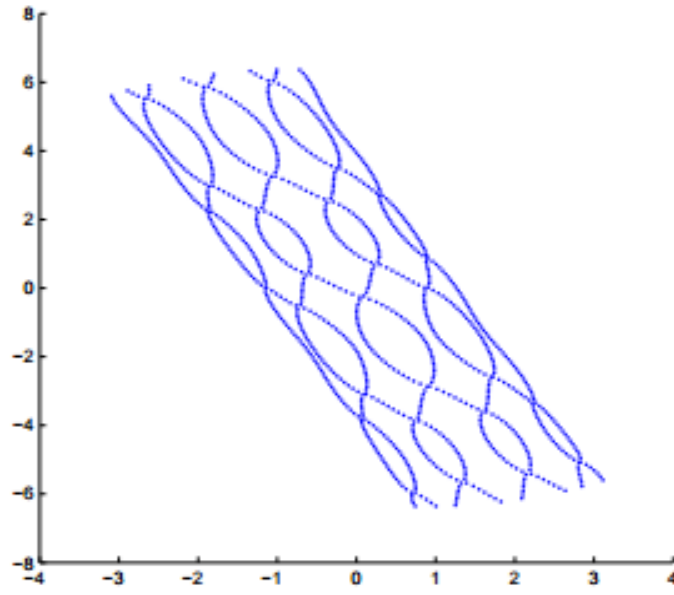
Orijinalde gösterilen fakat dönüştürülen datasette kaybolan yapı E hatası ile gösterilir ve aşağıdaki şekilde tanımlanır:

$$E = \frac{1}{\sum_{i < j} d_{ij}^*} \sum_{i < j}^n \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*} \quad (10)$$

Aslında hata farkların karesi alınarak bulunur fakat toplamaların $i < j$ aralığında olması ikili uzaklıkların bir kez sayılmış olması içindir.Topolojinin korunması paydada bulunan d_{ij}^* faktöründen kaynaklanır,bu faktör iki nokta arası orijinal uzaklığın küçük olmasından emin olunması içindir ve farkların karesinin ağırlıklandırılması daha büyüktür.

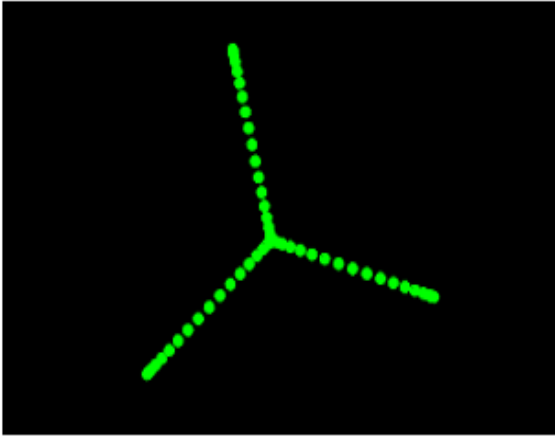
Sammon mapping in ana zayıf noktası uzaklık bulmaya çok daha fazla ağırlık vermesidir örnek olarak 10^{-5} yerine 10^{-4} gibi bir uzaklık bulmaktansa.Ana kullanım alanı görselleştirme amaçlarıdır[22]. Başarılı olarak gen datası[23] ve coğrafik dataların[24] uygulamalarında kullanılmıştır.

Arka sayfadaki Şekil 11'de,Şekil 9'da gösterilen üç boyutlu dataya Sammon mapping uygulandıktan sonra çıkan şekil verilmiştir.PCA'den farklı sonuç olarak daha az sinüzoidal kesişmesi vardır fakat topolojiyi PCA'e karşılık daha iyi korumuştur.

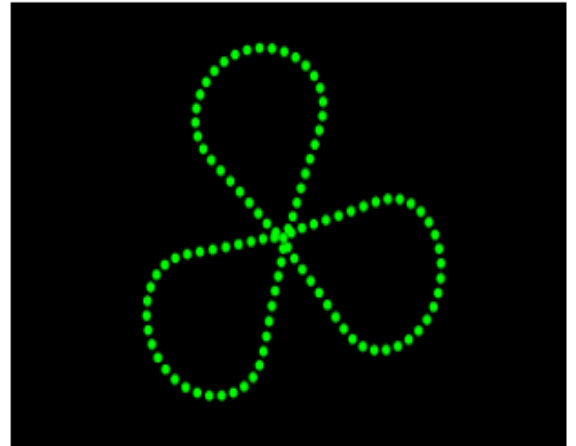


Şekil-11

Şekil 12 ve 13'de [25] de verilen altı boyutlu 'bouquet of circles' setinin PCA ve Sammon sonucunda iz düşümleri verilmiştir.



Şekil 12: PCA ile yapılan iz düşümde datasetin yapısı korunmamıştır.--Üç daire içerdiği belli değil.



Şekil 13: Sammon mapping topolojik yapıyı korumaktadır.--Daireler hasar görmüş olsa da hala tek noktada üç kapalı döngü görülmektedir

4.BULGULAR

Boyutluluk azaltmanın yüksek boyutlu bir uzayda gösterilen datanın uygulanan algoritmaya göre deęişen belli özelliklerin korunmasıyla birlikte düşük boyutlu bir uzayda izdüşümünü almak olarak söylenebilir.En çok kullanılan tekniklerden biri olan PCA lineer olarak başarılı olsa da gerçek dünya datası lineer deęildir ve buna göre kernel fonksiyonları kullanılarak Kernel PCA uygulanır.PCA uygulanınca non-lineer datanın yapısı korunamadığı için başka yöntemler kullanılır.(bkz. 3.5)

Korunması istenen özellikte teknięi olumlu veya olumsuz etkileyebilir MDS teknięi başarılı olsa da Öklid uzaklıkları kullanıldığı için eğrilmiş manifoldlarda başarısızlık göstermektedir ve bunun düzeltilmesi için Geodesic uzaklığın alındığı başka teknik(ler) kullanılabilir.(bkz 3.3)

5.TARTIŞMA VE SONUÇ

Boyutluluk azaltma tekniklerinin sınıflandırılması ve başlıca kullanılan bazı tekniklerin açıklamaları verilmiştir.Genele bakıldığında üç tane ana sorun vardır.

1. Amaç fonksiyonlarındaki sorunlar
2. Eigendecomposition'ların numerik zorlukları
3. Boyutluluğun lanetine olan duyarlılık

2 ve 3 numaralı eklerde python dili kullanılarak hazır datasetler üzerinde kullanılan tekniklerin sonucu gösterilmiştir.

Çalışmanın devamında teknikler incelenmeye devam edilecektir.

[]

KAYNAKLAR

- [1] L. Breiman. Random forests. Technical report, Department of Statistics, University of California, 2001.
- [2] Kovaryans, Wikipedia <https://tr.wikipedia.org/wiki/Kovaryans>
- [3] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572, 1901.
- [4] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.
- [5] T. Cox and M. Cox. *Multidimensional Scaling*. Chapman Hall, Boca Raton, 2nd edition, 2001.
- [6] M.A.Turk and A.P.Pentland. Facerecognition using eigenfaces. In *Proceedings of the Computer Vision and Pattern Recognition 1991*, pages 586–591, 1991
- [7] R. Huber, H. Ramoser, K. Mayer, H. Penz, and M. Rubik. Classification of coins using an eigenspace approach. *Pattern Recognition Letters*, 26(1):61–75, 2005
- [8] A.M. Posadas, F. Vidal, F. de Miguel, G. Alguacil, J. Pena, J.M. Ibanez, and J. Morales. Spatialtemporal analysis of a seismic series using the principal components method. *Journal of Geophysical Research*, 98(B2):1923–1932, 1993.
- [9] J.B. Tenenbaum, V. de Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- [10] E.W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.
- [11] R.W. Floyd. Algorithm 97: Shortest path. *Communications of the ACM*, 5(6):345, 1962.
- [12] M.Balasubramanian and E.L.Schwartz. The Isomap algorithm and topological stability. *Science*, 295(5552):7, 2002.
- [13] dimensional biomedical data sets by Isomap coordinates. In *Proceedings of the 16th IEEE Symposium on Computer-Based Medical Systems*, pages 50–55, 2003.
- [14] B. Raytchev, I. Yoda, and K. Sakaue. Head pose estimation by nonlinear manifold learning. In *Proceedings of the 17th ICPR*, pages 462–466, 2004.

- [15] M. Niskanen and O. Silvén. Comparison of dimensionality reduction methods for wood surface inspection. In *Proceedings of the 6th International Conference on Quality Control by Artificial Vision*, pages 178–188, Gatlinburg, TN, USA, 2003. International Society for Optical Engineering
- [16] B. Schölkopf, A. J. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998.
- [17] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, Cambridge, UK, 2004
- [18] K.I. Kim, K. Jung, and H.J. Kim. Face recognition using kernel principal component analysis. *IEEE Signal Processing Letters*, 9(2):40–42, 2002.
- [19] A. Lima, H. Zen, Y. Nankaku, C. Miyajima, K. Tokuda, and T. Kitamura. On the use of Kernel PCA for feature extraction in speech recognition. *IEICE Transactions on Information Systems*, E87-D(12):2802–2811, 2004.
- [20] H. Hoffmann. Kernel PCA for novelty detection. *Pattern Recognition*, 40(3):863–874, 2007.
- [21] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Prentice Hall, 1988.
- [22] M. Martin-Merino and A. Munoz. A new Sammon algorithm for sparse data visualization. In *Proceedings of the 17th International Conference on Pattern Recognition*, pages 477–481, 2004.
- [23] S. Ekins, K.V. Balakin, N. Savchuk, and Y. Ivanenkov. Insights for human ether-a-go-go-related gene potassium channel inhibition using recursive partitioning and Kohonen and Sammon mapping techniques. *Journal of Medicinal Chemistry*, 49(17):5059–5071, 2006.
- [24] M. Takatsuka. An application of the self-organizing map and interactive 3-D visualization to geospatial data. In *Proceedings of the 6th International Conference on GeoComputation*, 2001. |

EKLER

1. <http://setosa.io/ev/principal-component-analysis/>
2. http://scikit-learn.org/stable/auto_examples/manifold/plot_manifold_sphere.html
3. http://scikit-learn.org/stable/auto_examples/manifold/plot_compare_methods.html#sphx-glr-auto-examples-manifold-plot-compare-methods-py