

# Bölüm 12

## LOKAL POLİNOMAL REGRESYON (LPR) VE DÜZGÜNLEŞTİRME

*H.Nazan ÇAĞLAR<sup>1</sup>*



## 1. Giriş

İstatistikte verilerin modellenmesinde regresyon analizi önemli bir yere sahiptir. Verilere ait regresyon modelinin bulunmasında bir çok matematiksel fonksiyon kullanılır. Uygulamalı istatistikte düzgünleştirilmiş fonksiyonların elde edilmesi önemli bir alandır. Bu amaçla kullanılan bir çok popüler yöntem vardır. Bunların en çok kullanılanları smoothing spline, kernel smoothers ve lokal polinomal regresyondur (LPR) . Lokal polinomlar , parametrik olmayan regresyon modelleri için de önemli bir yere sahiptir. Uygun optimal ağırlık fonksiyonlarının seçimiyle smoothing işlemlerinde kullanılır. ( Seifert & Gasser , 1996) Lokal polinomların band seçimi üzerine yapılmış bir çok çalışma vardır . (Linton et al., 2001 ) .

En iyi bilinen parametrik olmayan regresyon tahminçileri Nadarya-Watson ve Gasser-Muller Kernel tahminçileri, Smoothing spline ve Lokal polinomlardır. Bu polinomların özellikleri ( Simonoff ,1996 ) ayrıntılı incelenmektedir. Zaman serileri, Çapraz tabloların ( Contingency table) düzgünleştirme işlemlerinde Lokal polinomlar kullanılmaktadır. Ayrıca matematik ,fizik, biyoloji gibi bilimsel alanlarda özellikle diferansiyel denklemlerin sayısal çözümlerinde LPR yöntemi etkin olarak kullanılmaktadır. Bu çalışmada LPR nin etkinliği farklı alanlardan alınan örnekler üzerinde incelenmiştir.

## 2. LOKAL POLİNOMAL REGRESYON

$(x_1, y_1), \dots, (x_n, y_n)$  gözlemleri için aşağıdaki fonksiyonel modeli düşünelim.

$$y_i = f(x_i) + \varepsilon_i$$

Burada  $x_i$  bağımsız değişken,  $y_i$  ler bağımlı değişken ve  $\varepsilon_i$  hata terimidir.

Taylor seri açılım yaklaşımı ile  $x_i$  noktası civarında  $f(x)$  fonksiyonunu seriye açalım,

$$f(x_i) = f(x) + f'(x)(x_i - x) + \dots + \frac{f''(x)(x_i - x)^2}{2!}$$

$f(x)$  fonksiyonunun bulunması lokal polinomal regresyon olarak bilinir. Kernel yaklaşımı kullanılarak aşağıdaki denklem en küçük kareler optimizasyonu ile çözülür.

$$\min(\beta) \sum_{i=1}^n [y_i - \beta_0 - \beta_1(x_i - x) - \beta_2(x_i - x)^2 - \dots - \beta_p(x_i - x)^p]^2 K\left(\frac{x_i - x}{h}\right) \quad (1)$$

Lokal polinomların  $\beta_j$  tahmincileri yukarıda verilen toplamın minimize edilmesi sonucunda elde edilir.

$\beta = (\beta_0, \dots, \beta_p)'$  tahmin edilecek parametrelerdir.

$x_1$  : 0 ile 1 arasında bir nokta ,

$K$  : negatif olmayan ağırlık fonksiyonu ( Kernel Function )

$h$  : band genişliği

$x_0$  : Lokal polinom içinde herhangi bir nokta ( orta nokta )

$$X = \begin{bmatrix} 1 & (x_1 - x_1) & \dots & (x_1 - x_1)^p \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x_1) & \dots & (x_n - x_1)^p \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}, \quad W = \text{diag}\left(K\left(\frac{x_i - x_0}{h}\right)\right)$$

$$\hat{\beta} = (X' W X)^{-1} X' W Y \quad (2)$$

Olarak bulunur. Sonuç olarak Lokal polinomal yaklaşımda  $y$  tahmin değerleri

$$\hat{y} = \sum_{j=0}^p \hat{\beta}_j (x - x_1)^j \quad (3)$$

formülünden hesaplanır.

Kernel fonksiyonu ( Yang & Yue , Nov.2014) olarak Tablo1 de verilen fonksiyonlar seçilebilir. Çalışmalarda Epanechnikov Kernel fonksiyonu kullanılmıştır. Polinomun derecesi tahminler için önemli bir

faktördür. Polinom derecesi 3 den büyük olduğunda tahminlerdeki yanlışlık azalmaktadır. Tek dereceden polinomlar daha etkin sonuçlar vermektedir. Kernel seçimi ve bant genişlikleri (h) de tahminler için bir diğer faktördür.

**Tablo 1. Kernel Fonksiyonları**

Kernel	K(t)
Epanechnikov	$\frac{3}{4} \left(1 - \frac{1}{5}t^2\right) / \sqrt{5}, \quad  t  < \sqrt{5}$
	0                      diğer durumlar
Biweight	$\begin{cases} \frac{15}{16} (1-t^2)^2, &  t  < 1 \\ 0 & \text{diğer durumlar} \end{cases}$
Triangular	$\begin{cases} 1- t , &  t  < 1 \\ 0 & \text{diğer durumlar} \end{cases}$
Gaussian	$\frac{1}{\sqrt{2\pi}} e^{-(1/2)t^2}$
Rectangular	$\begin{cases} 1/2, &  t  < 1 \\ 0 & \text{diğer durumlar} \end{cases}$

**Problem 2.1.** Aşağıdaki gözlem değerleri için farklı polinom derecelerinde LPR modeli kurularak hata kareleri toplamı ( HKT =  $\sum (y_i - \hat{y}_i)^2$  ) tablo 2 de verilmektedir. Ayrıca polinom derecesi 3 için LPR modelin grafiği Şekil 1 de sunulmaktadır. LPR metodu aşağıdaki datalar için Matlab kullanılarak uygulanmıştır ve Matlab kodu Ek1 de paylaşılmıştır. Tablo 2 den polinom derecesi yükseldikçe HKT düştüğü görülmektedir bu da yöntemin etkinliğini göstermektedir. Şekil 1 den metodun etkinliği görülmektedir.

Gözlem	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x	2	5	7	10	14	19	26	31	34	38	45	52	53	60	65
y	54	50	45	37	35	25	20	16	18	13	8	11	8	4	6

Datalar için oluşturulan çözüm matrisleri aşağıda verilmektedir.

$$Y = [54 \ 50 \ 45 \ 37 \ 35 \ 25 \ 20 \ 16 \ 18 \ 13 \ 8 \ 11 \ 8 \ 4 \ 6]'$$

$$W = \text{diag} (0,32)$$

$$X = \begin{bmatrix} 1 & -5,32 & 28,30 & -150,53 \\ 1 & -2,32 & 5,38 & -12,48 \\ 1 & -0,32 & 0,10 & -0,03 \\ 1 & 2,68 & 7,18 & 19,26 \\ 1 & 6,68 & 44,63 & 298,14 \\ 1 & 11,68 & 136,43 & 1593,59 \\ 1 & 18,68 & 348,96 & 6518,71 \\ 1 & 23,68 & 560,76 & 13279,12 \\ 1 & 26,68 & 711,85 & 18992,37 \\ 1 & 30,68 & 941,29 & 28879,18 \\ 1 & 37,68 & 1419,82 & 53499,29 \\ 1 & 44,68 & 1996,34 & 89197,44 \\ 1 & 45,68 & 2086,70 & 95321,51 \\ 1 & 52,68 & 2775,23 & 146200,29 \\ 1 & 57,68 & 3327,03 & 191904,76 \end{bmatrix}$$

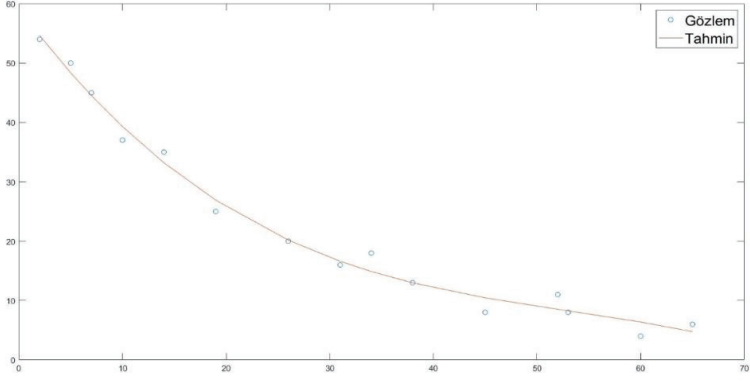
$\beta_j$  tahminçileri ise  $\hat{\beta} = [43.9389 \ -1.824 \ 0.0342 \ -0.0002]'$  olarak hesaplanır. Y tahmin değerleri ise

$$\hat{y} = [54.65 \ 48.36 \ 44.52 \ 39.29 \ 33.2 \ 26.9 \ 20.18 \ 16.62 \ 14.89 \ 12.98 \ 10.46 \ 8.52 \ 8.27 \ 6.38 \ 4.76]'$$

Olarak hesaplanır.

**Tablo 2.** Polinom derecesine bağlı HKT

Derece	HKT
2	72.34
3	44.95
4	43.74
5	41.73

**Şekil 1.** Polinom derecesi üç için model grafiği

### 3. İKİ BOYUTLU TABLOLARDA ( crosstabs ) LOKAL POLİNOMAL YAKLAŞIMLA TAHMİN

RxC boyutlu bir tabloda hücrelerin düzgünleştirilmiş (smooth ) frekansları lokal polinomal yaklaşımla tahmin edilebilir. Tablo 2 deki gözlem değerleri için sözü edilen yöntemle düzgünleştirilmiş tahminler, iki boyutlu tabloyu daha iyi açıklayıcı duruma getirir.

Yazılan bir bilgisayar programı ile literatürden ( Simonoff ,1996 ) alınmış problem için lokal polinomal yaklaşımın iki boyutlu tablolar için uygulaması aşağıda açıklanmaktadır. Denklem 2 deki W matrisi çok değişkenli ( problem 3.1 de 2 değişken kullanıldı) Kernel fonksiyonu kullanılarak bulunur.

Çapraz tablonun satır ve sütun numaraları X açıklayıcı değişken matrisini oluşturur. Tablonun hücre değerleri bağımlı değişken Y matrisini oluşturmaktadır.

**Problem 3.1.** 55 MBA öğrencisinin, İşletme eğitiminde, İstatistik ve Ekonomi nin önemi hakkında sorulara verdikleri yanıtlar Tablo 3 de görülmektedir. Bu tabloda satırlar **İstatistik** için verilen yanıtları, sütunlar ise **ekonomi** için verilen yanıtları içermektedir. Yanıtlar şu şekilde kodlanmıştır;

1= tamamen gereksiz , 2 = gereksiz , 3= az önemli , 4= tarafsız , 5=önemli , 6= çok önemli , 7= kesinlikle önemli

**Tablo 3.** MBA öğrencilerinin cevaplarına ait çapraz tablo

		Ekonomi						
		1	2	3	4	5	6	7
İstatistik	2	0	1	0	0	0	1	0
	3	0	0	0	1	0	0	0
	4	0	0	3	6	4	0	0
	5	0	0	1	4	7	4	0
	6	1	0	0	2	6	10	1
	7	0	0	0	0	0	2	1

Bu data'lara ait yoğunluğa bağlı renklendirilmiş grafik şekil 2 de görülmektedir. Tablo 3 verilerine lokal polinomal regresyon uygulanarak düzleştirilmiş yoğunluklar şekil 3 de görülür. Ekonomi ve İstatistik'in işletme eğitiminde oldukça önemli olduğu ( koyu renkli kısımdan ) görülmektedir.

0	1	0	0	0	1	0
0	0	0	1	0	0	0
0	0	3	6	4	0	0
0	0	1	4	7	4	0
1	0	0	2	6	10	1
0	0	0	0	0	2	1

Şekil 2. Tablo 3 datalara ait grafik

-0,2	0,1	0,4	0,7	1,0	1,3	1,6
0,0	0,3	0,6	0,9	1,2	1,5	1,8
0,3	0,6	0,9	1,2	1,5	1,8	2,1
0,6	0,8	1,1	1,4	1,7	2,0	2,3
0,8	1,1	1,4	1,7	2,0	2,3	2,5
1,1	1,4	1,6	1,9	2,2	2,5	2,8

Şekil 3. Tablo 3 datalarına LPR uygulanmış grafiği

Örnekten , yorumlanması zorluk gösteren tabloların LPR ile daha iyi açıklanabilir hale geldiği görülmektedir .

#### 4. ZAMAN SERİLERİNİN LOKAL POLİNOMAL YAKLAŞIMLA DÜZGÜNLEŞTİRİLMESİ

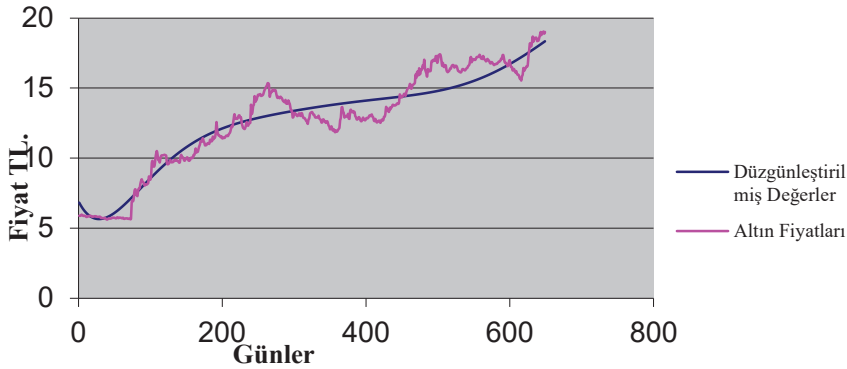
Zaman serilerinin modellenmesinde özellikle stokastik trendin bulunması büyük önem taşır. Bu amaçla kullanılan bir çok yöntem vardır. Stokastik trendin elde edilmesinde ,düzgünleştirme problemlerinde kullanılacak alternatif bir yöntemde Lokal Polinomal yaklaşımdır. Özellikle son yıllarda bu alanda yapılan bilimsel çalışmalarda Kernel fonksiyonları çok geniş yer tutmaktadır.

##### Problem 4.1.

Bu bölümde stokastik trend bulmak için Türkiye’deki altın fiyatları kullanılarak lokal polinomal yaklaşım kullanılmıştır (Çağlar vd, 2004) . Analizde 678 gözlem kullanılmıştır. Formül (3) deki  $\hat{\beta}$  değerleri Matlab programı yardımıyla bulunmuştur ve Matlab kodu Ek1 de paylaşılmıştır

grafik olarak Şekil 4 de altın fiyatlarının düzgünleştirilmiş eğrisi görülmektedir.

Lokal polinomal yaklaşımla bulunan zaman serisinin stokastik trendi şekilden görülmektedir. Düzgünleştirilmiş stokastik trendin elde edilmesinde local polinomal yaklaşımın etkin bir yöntem olduğu söylenebilir. Farklı Kernel'ler seçilerek MSE ( Mean Square Error ) ler karşılaştırılarak optimum kernel seçimi yapılabilir.

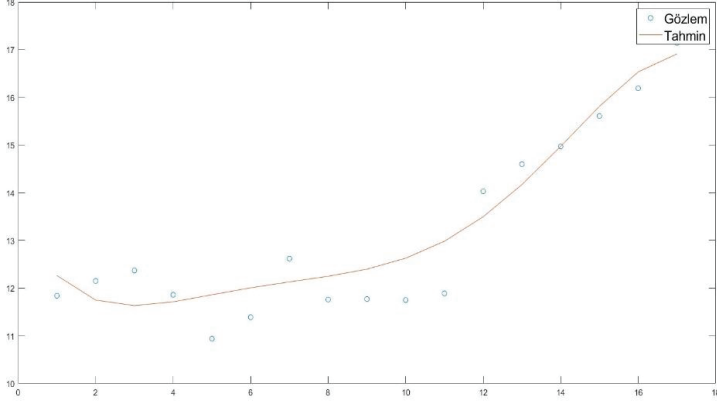


Şekil 4. Altın fiyatlarının değişim grafiği ( düzgünleştirme öncesi ve sonrası)

**Problem 4.2.** Tüik den alınan 2019 –2021 yılları arasındaki aylık TÜFE oranları kullanılarak polinom derecesi beş için LPR ile bulunan stokastik trend Şekil 5 de görülmektedir. Matlab kodu Ek1 de paylaşılmıştır. Ayrıca Tablo 4 de HKT farklı polinom dereceleri için verilmiştir. Tablo 4 den polinom derecesi yükseldikçe HKT düştüğü görülmektedir bu da yöntemin etkinliğinin bir kez daha görülmesini sağlamaktadır.

Tablo 4. Polinom derecesine bağlı HKT

Derece	HKT
2	4.473
3	4.463
4	3.619
5	3.618



Şekil 5. Tüfe aylık oranlarının değişim grafiği

## 5. LPR 'nin İNTEGRAL DENKLEMLERE UYGULANMASI

Bu bölümde, demografi , malzeme, sigortacılık alanlarında birçok uygulaması bulunan Volterra integral denklemini ( Delves & Mohamed, 1985) LPR yaklaşımı ile çözelim.

Denklemin genel formu aşağıda verilmektedir.

$$x(t) + \lambda \int_a^t k(t, s)x(s)ds = y(t), \quad a \leq t \leq b, a \leq s \leq b \quad (4)$$

Denklem (3) denklem (4) ün yaklaşık çözümü olsun.

$$\sum_{j=0}^p \beta_j (t - t_0)^j + \lambda \sum_{j=0}^p \beta_j \int_a^t k(t, s)(s - t_0)^j ds = y(t),$$

$$a \leq t \leq b, a \leq s \leq b \quad (5)$$

matris formunda yazılırsa,

$$X = \begin{bmatrix} 1 & (t_1 - t_0) + \lambda \int_a^{t_1} k(t, s)(s - x_0) ds & \dots & (t_1 - t_0)^p + \lambda \int_a^{t_1} k(t, s)(s - x_0)^p ds \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & (t_n - t_0) + \lambda \int_a^{t_n} k(t, s)(s - x_0) ds & \dots & (t_n - t_0)^p + \lambda \int_a^{t_n} k(t, s)(s - x_0)^p ds \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

$$K(u) = \begin{cases} \frac{15}{16}(1 - u^2)^2 & \text{if } |u| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Kernel fonksiyonu olarak quartic kernel kullanılmıştır.

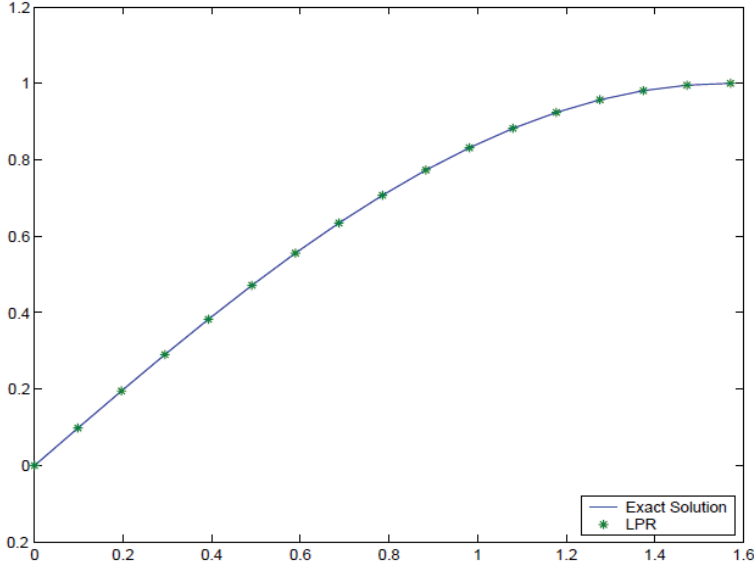
(6) denklem (2) de yerine yazılırsa  $\beta_i$  değerleri matris sistemi çözümlenerek elde edilir. Böylece verilen denklemin yaklaşık çözümü elde edilmiş olur.

### Problem 5.1.

Aşağıda verilen integral denklemin (Çağlar vd, 2008 ) LPR ile sayısal çözümünü bulalım.

$$x(s) = \sin(s) - s + \int_0^{\frac{\pi}{2}} stx(t)dt, \quad 0 \leq t \leq \frac{\pi}{2}$$

Denklemin teorik çözümü  $x(s) = \sin(s)$  olarak bilinmektedir. Sayısal çözüm değerleri Tablo 5 de verilmekte ve Şekil 6 de gösterilmektedir. LPR yöntemin integral denklemini çözmekte iyi bir yakınsaklık sağladığı görülmektedir. Yöntem aynı zamanda farklı bant genişliği ve Kernel fonksiyonları için denenebilir.

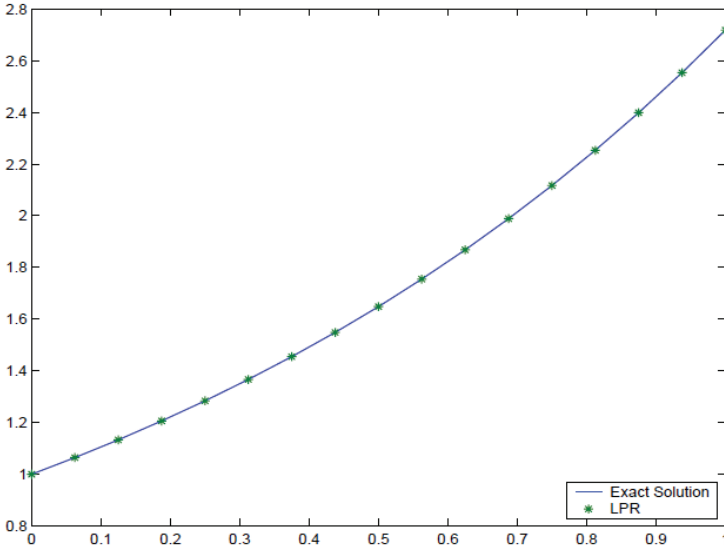
Şekil 6.  $x(s)=\sin(s)$  ve LPR çözümüTablo 5.  $y_i$  teorik çözüm,  $Y_i$  LPR çözümü

$y_i$	$Y_i$
0	-0.00000000120664
0.09801714032956	0.09801716335287
0.19509032201613	0.19509030895861
0.29028467725446	0.29028466194663
0.38268343236509	0.38268343573862
0.47139673682600	0.47139675044662
0.55557023301960	0.55557023796561
0.63439328416365	0.63439327155392
0.70710678118655	0.70710675890148
0.77301045336274	0.77301043768579
0.83146961230255	0.83146961261548
0.88192126434835	0.88192127296150
0.92387953251129	0.92387952957580
0.95694033573221	0.95694031039742
0.98078528040323	0.98078525344602
0.99518472667220	0.99518473630261
1.00000000000000	0.9999998107762

### Problem 5.2.

Teorik çözümü  $x(s) = e^s$  olan aşağıdaki denkleminin çözümü grafik olarak Şekil 7 de görülmektedir. Şekilden teorik çözüme yakınsama çok net görülmektedir.

$$x(s) = e^s - \frac{e^{s+1}-1}{s+1} + \int_0^1 e^{st}x(t)dt, 0 \leq t \leq 1$$



Şekil 7.  $x(s) = e^s$  ve LPR çözümü

## 6. SONUÇ

Local polinomal regresyon yöntemi , mühendislik, istatistik ve matematik alanlarında yaygın kullanılmaktadır. Yaklaşım fonksiyonun bulunmasında en küçük kareler yöntemi uygulama kolaylığına sahip olmakla birlikte yöntemin kodları Matlab veya farklı bir proglama dilinde kolaylıkla yazılabilir. Çalışmada Lokal polinomların teorisi anlatıldıktan sonra çapraz tablolar, zaman serileri ve mühendislik denklemlerinde uygulamaları ele alınmıştır. Farklı polinom derceleri , Kernel ve bant seçimlerine bağlı olarak optimal düzgünleştirme sağlanabildiği görülmüştür. Sonuç olarak , LPR nin farklı alanlardaki problemlerin sayısal çözümlerinin elde edilmesinde , düzgünleştirme işlemlerinde ve çapraz tablolarda yoğunluk alanlarının belirlenmesinde oldukça etkili bir yöntem olduğu görülmüştür.

## KAYNAKLAR

- Çağlar H. ,Çağlar N. (2008). Numerical solution of integral equations by using local polynomial regression. *Journal of Computational Analysis and Applications* 10(2), 187-195.
- Çağlar N ,Çağlar H. (2004). Stokastik trendin lokal polinomal yaklaşım ve üstel düzgünleştirme ile bulunmasında hata kareleri ortalamasının karşılaştırılması. *Öneri*, 6(22), 301-304.
- Delves, L.M. and Mohamed,J.L.(1985) *Computational Methods for Integral Equations*. Cambridge University Press.
- Linton O.,Mammen E., Nielsen J,Tanggaard C. (Nov 2001) Yield curve estimation by Kernel smoothing methods. *Journal of Econometrics*, 105(1).
- Seifert B. And Gasser T. (March 1996) Finite-Sample variance of local polinomials : analysis and solutions. *Journal of American Statistical Association*. 91(433).
- Simonoff J.S. *Smoothing Methods in Statistics* . (1996) Springer-Verlag New York.
- Yang F.,Yue Z. (Nov.2014). Kernel density estimation of three-parameter Weibull distribution with neural network and genetic algorithm. *Applied Mathematics and computation*, 247,803-814

### İnternet Kaynakları

- 1-<https://data.tuik.gov.tr/Kategori/GetKategori?p=Enflasyon-ve-Fiyat-106>, 2021
- 2- [http://www.cnachtsheim-text.csom.umn.edu/kut86916\\_ch13.pdf](http://www.cnachtsheim-text.csom.umn.edu/kut86916_ch13.pdf)
- 3- Pusula Yazılım , Donanım ve Danışmanlık Hizmetleri Ltd. Şti.

**Ek1. Matlab Kodu**

```

function [s,x,xk,b,ys] = lpr(x1,y,p,k,m)

    [n,n1]=size(x1);x=zeros(n,p+1);
    x(:,1)=ones(n,1);
    h=(x1(n)-x(1))/m;

    for i=1:n    t1=0; t2=x1(1);

    for j=x1(1):h:x1(n)

        if x1(i) < j    x0 = j - h / 2; break;

        end;
    end;

    xk1=kernel((x(i) - x0) / h,k);    xk(i,i)=xk1;

    t1 = t1 + xk1; t2 = t2 + xk1 * x(i);
    end;

    x3=x1-t2 / t1;

    for i=0:p    x(:,i+1)= x3.^i; end

    b=inv(x'*xk*x)*x'*xk*y;

    ys=x*b;

    figure
    set(gca,'FontSize',45)
    plot(x1,y,'o',x1,ys);
    k=legend('Gözlem','Tahmin')
    k.FontSize = 16;

    s=sum((y-ys).^2)

end

```