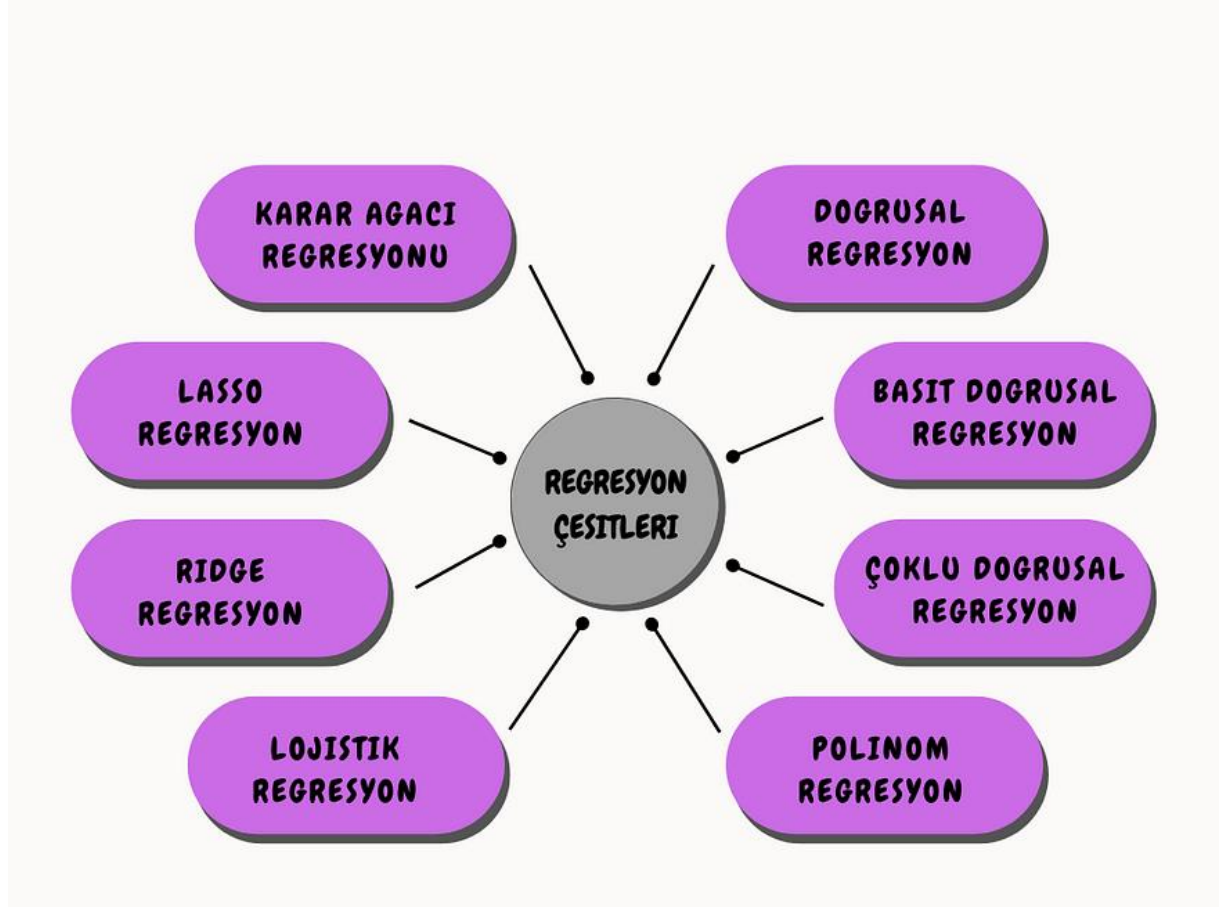


## Regresyon eřitleri:



### 1-Doğrusal Regresyon:

Doğrusal regresyon, iki deėiřken arasındaki doğrusal iliřkinin bir doğru denklemini olarak tanımlanıp, deėiřkenin deėerlerinden biri bilindiėinde diėeri hakkında tahmin yapılmasını saėlar.

İř arkadaşlarınızın daha önce zaten anladıklarını görmüş ve düşünmüş olabilecekleri kalıpları ve iliřkileri ortaya ıkararak daha iyi iç görüler saėlamak için de doğrusal regresyondan yararlanabilirsiniz. Örneėin, satış ve satın alma verilerini analiz etmek, belirli günlerdeki ya da belirli saatlerdeki belirli satın alma kalıplarını ortaya ıkarmanıza yardımcı

olabilir. Regresyon analizinden toplanan iç görüler, iş liderlerinin şirketlerinin ürünlerinin yüksek talep göreceği zamanları tahmin etmelerine yardımcı olabilir.

Doğrusal regresyon, basit doğrusal regresyon ve çoklu doğrusal regresyon olarak iki başlık altında incelenir.

## **Basit doğrusal regresyon:**

Yanıt değişkeni ile tek bir açıklayıcı değişken arasındaki doğrusal ilişkiyi açıklar. Eğer tek bir yanıt değişkeni ve birden fazla açıklayıcı değişken arasındaki doğrusal veya eğrisel bir ilişki tanımlanmak istenirse, ilişki *çoklu doğrusal regresyon* ile incelenir .

Basit doğrusal regresyonda,  $Y$  yanıt değişkeni,  $X_1$  açıklayıcı değişkeni,  $\beta_0$  ve  $\beta_1$  bu değişkenin bilinmeyen parametrelerini ve  $\varepsilon_i$  şansa bağlı hata terimlerini ifade etmek üzere basit doğrusal regresyon modeli;

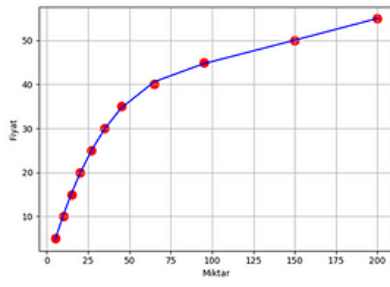
$$Y = \beta_0 + \beta_1 X_{1i} + \varepsilon_i \quad i = 1, 2, \dots, n$$

olarak yazılabilmektedir (Kutner ve ark., 2005). Çoklu doğrusal regresyon modeli,  $p$  adet açıklayıcı değişken ve  $n$  adet gözlem için;

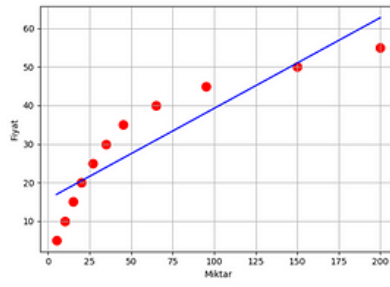
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i \quad i = 1, 2, \dots, n$$

şeklinde yazılabilmektedir (Kutner ve ark., 2005; Weisberg, 2005).

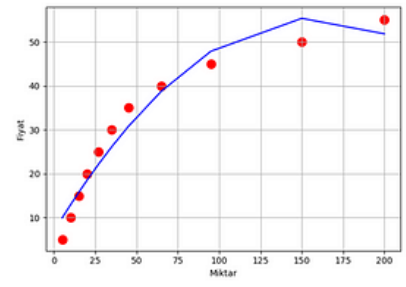
## **2-Polinom Regresyon:**



Overfit  
Yüksek Varyans  
Düşük Bias



Underfit  
Düşük Varyans  
Yüksek Bias



Correct Fit  
Düşük Varyans  
Düşük Bias

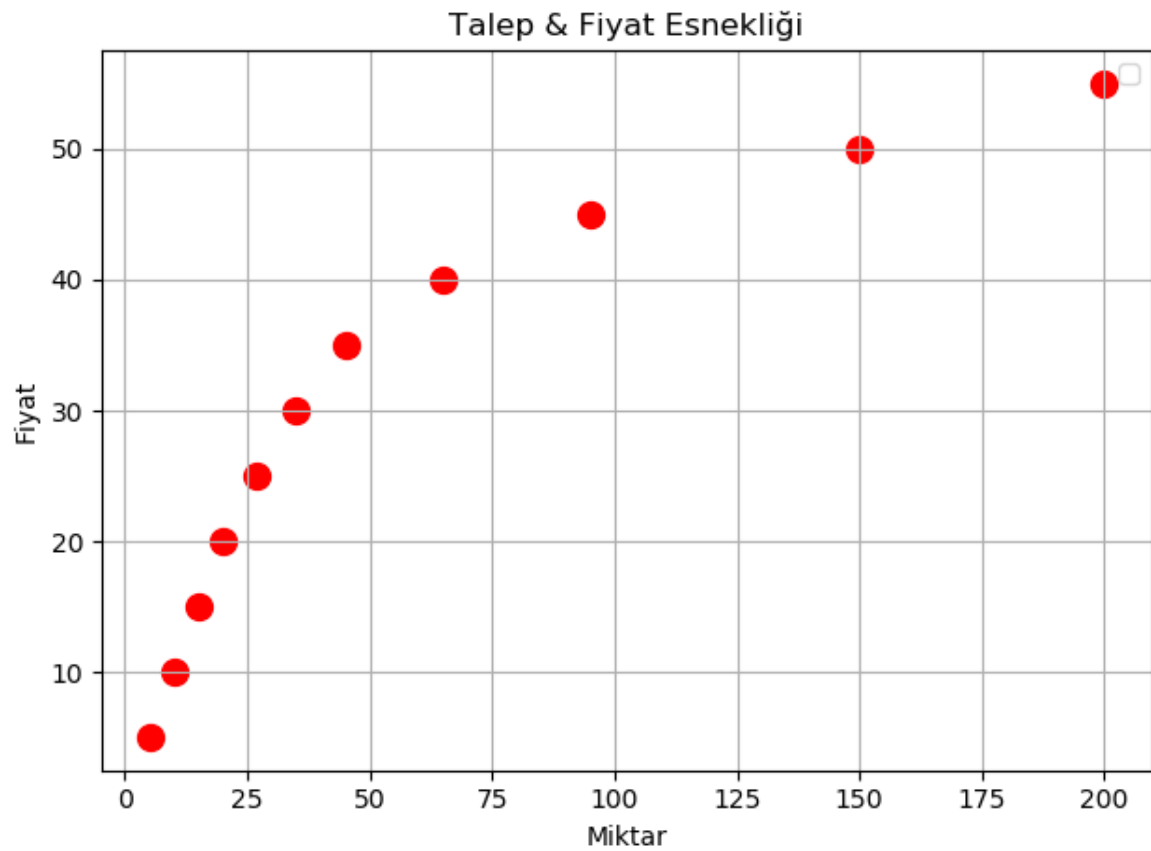
Doğrusal bir regresyonda girdi (x) ve çıktı (y) arasında lineer (doğrusal/line) bir ilişki olduğu varsayılmaktadır. Öte yandan polinomsal regresyonda ise girdi ve çıktı arasındaki ilişki düz bir doğru biçiminde değildir. Burada bir doğru yerine eğriden (curve) bahsedebiliriz. Ve temelde polinomsal regresyon bize bu eğrinin fonksiyonunu vermektedir. Burada önemli olan kısım bu polinom fonksiyonunun hangi dereceyi alacağıdır (nth-degree polynomial).

## Neden Polinomsal Regresyon?

Aşağıdaki Python kodunda yer alan örnekte; talep edilen miktar ve fiyat noktalarının birleşimi scatter plot üzerinden gösterilmiştir (Price Elasticity). Bu veriler üzerinden yola çıkarak basit doğrusal regresyon ile polinomsal regresyonu karşılaştırıp inceleyelim.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt# talep miktarı: x
quantity = [5,10,15,20,27,35,45,65,95,150,200]# fiyat: y
price = [5,10,15,20,25,30,35,40,45,50,55]# Dataframe objesi
```

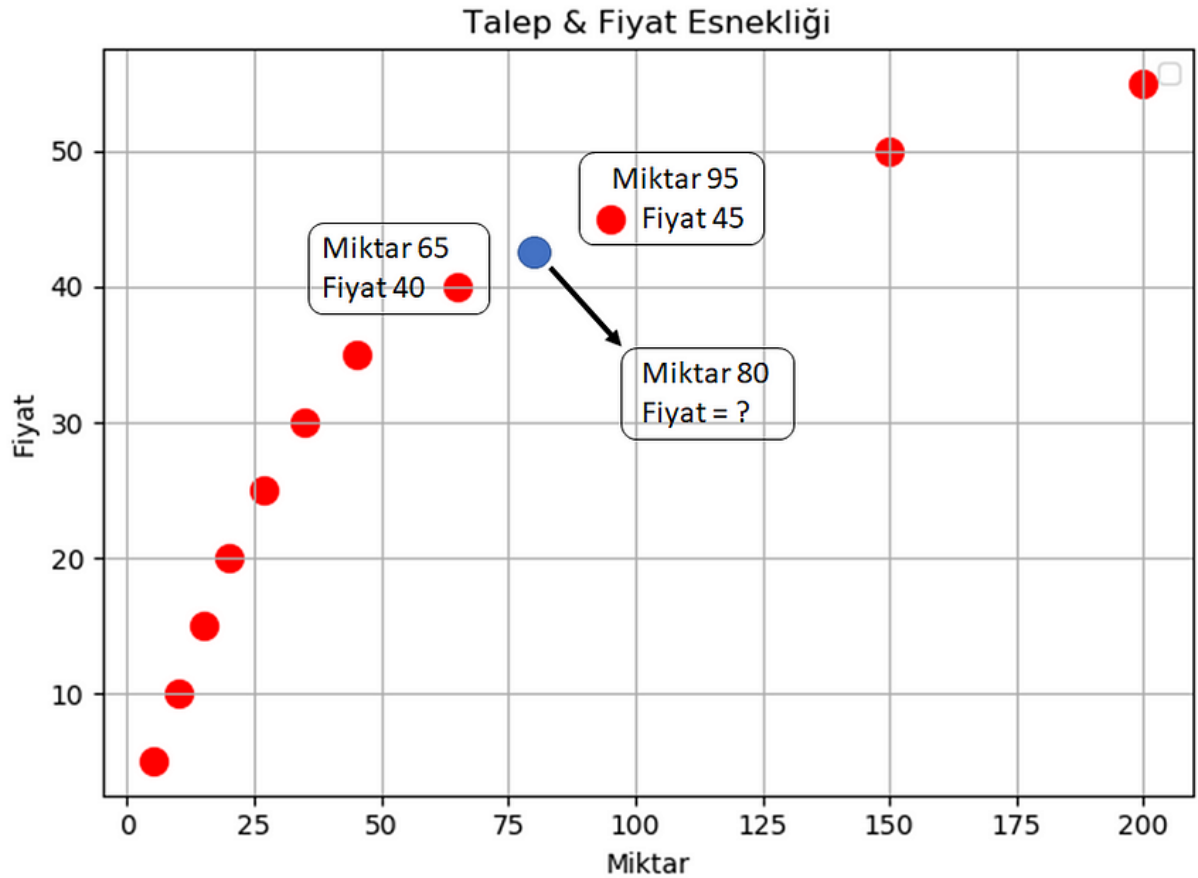
```
df = pd.DataFrame({"quantity":quantity,  
                  "price":price})# scatter plot gösterimi  
plt.scatter(df["quantity"],  
            df["price"],  
            s=100,  
            c="red",  
            edgecolors='red'  
            )plt.title("Talep & Fiyat Esnekliği")  
plt.xlabel("Quantity")  
plt.ylabel("Price")  
plt.grid(True)  
plt.legend()  
plt.show()
```



Yukarıdaki scriptte görüldüğü üzere önce gerekli olan kütüphaneler verilmiştir. Ardından iki adet örnek miktar/fiyat listesi oluşturularak bunlar DataFrame yapısına alınmıştır. Ardından talep edilen miktar ve fiyat değerlerinin kesişim noktaları için scatter plot çizilmiştir.

El ile çözebileceğimiz bir problemde yola çıkarsak polinomsal regresyonun temel mantığını anlamaya çalışalım.

Örneğin talep edilen *miktar* 80 ise *fiyatı* kaç bulabiliriz diye sorduğumuzda göz kararı bulunduğu noktaya en yakın iki noktayı inceleyerek buradaki fiyatların ortalamasını alabiliriz.



Fiyatını hesaplamak istediğimiz miktar olan 80'e en yakın iki değerin miktar ve fiyatlarını yukarıdaki scatter plot üzerinden de görebilirsiniz. Bu iki değerin fiyatlarının ortalamasını alalım.

$$y = (40 + 45) / 2 = 42,5$$

### 3– Karar Ağacı Regresyonu:

Karar Ağaçları Algoritmaları hem sınıflandırma da hem de regresyonda kullanılır. Regresyon için kullanılan algoritmayı şöyle açıklayabiliriz; Bağımsız değişkenleri bilgi kazancına göre aralıklara ayırır. Tahmin esnasında bu aralıktan bir değer sorulduğunda cevap olarak bu aralıktaki (eğitim esnasında öğrendiği) ortalamayı verir. Belli aralıklarda istenilen değerler için aynı sonuçları ürettiğinden kesikli bir modeldir.

### 4-Lojistik Regresyon:

Lojistik regresyon, istatistikte kullanılan bir model oluşturma tekniği olup iki ya da daha fazla sınıfta ifade edilebilen kesikli verilerde yanıt değişkeni (Y) için bir model oluşturma tekniğidir.

Örneğin, web sitesi ziyaretçinizin alışveriş sepetindeki ödeme düğmesine tıklayıp tıklamayacağını tahmin etmek istediğinizi varsayalım. Lojistik regresyon analizi, web sitesinde harcanan zaman ve sepetteki ürün sayısı gibi geçmiş ziyaretçi davranışlarına bakar. Geçmişte, ziyaretçiler sitede beş dakikadan fazla zaman geçirdiyse ve sepete üçten fazla ürün eklediysen ödeme düğmesine tıkladıklarını belirler. Lojistik regresyon işlevi bu bilgiyi kullanarak daha sonra yeni bir web sitesi ziyaretçisinin davranışını tahmin edebilir.

*Ridge Regresyon, Lasso Regresyon, Sıralı Regresyon gibi daha bir çok regresyon türü bulunmaktadır.*

