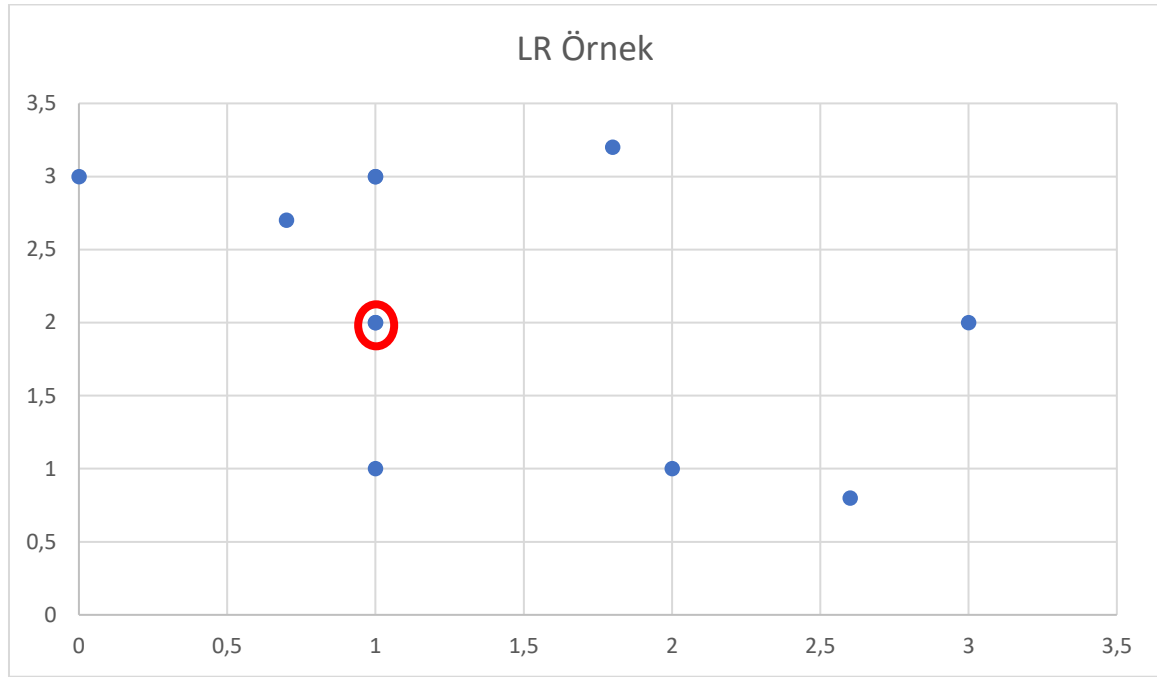
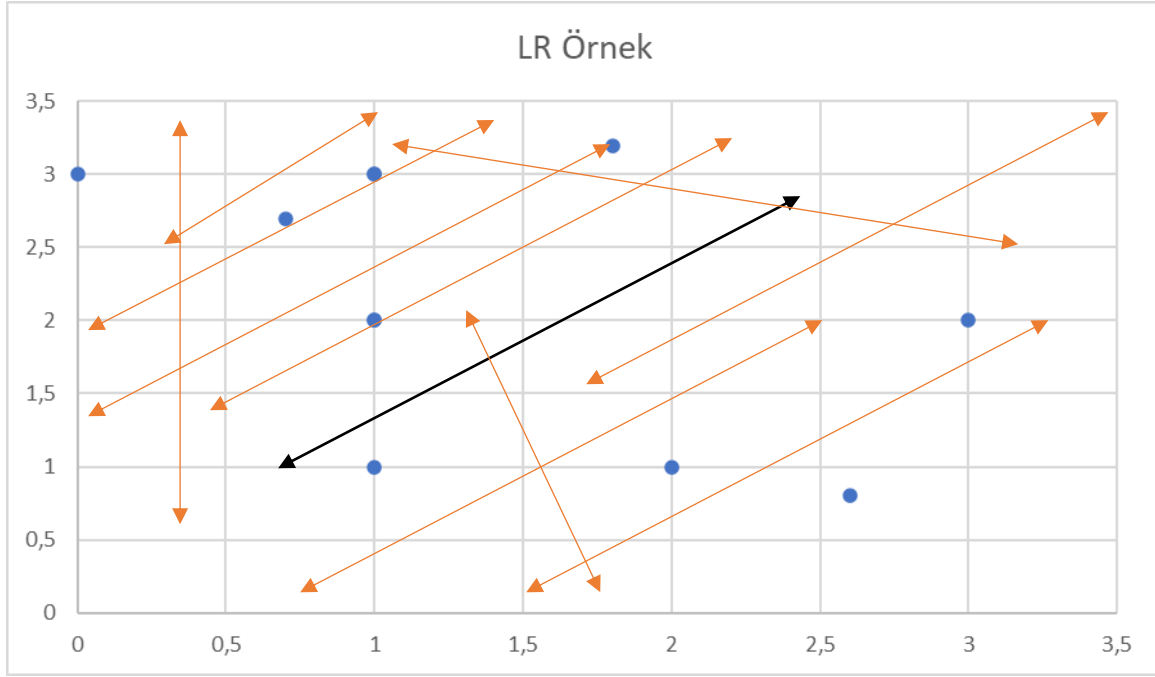


Lineer Regresyon Nedir?

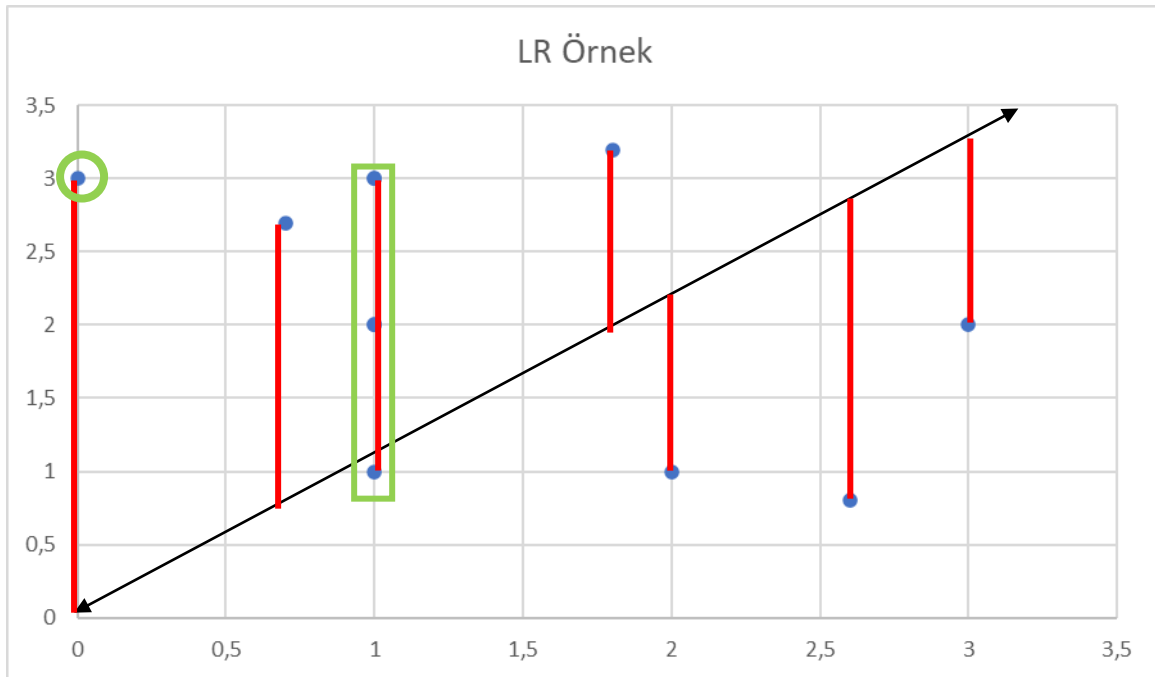
‘Python ile Makine Öğrenimi’ yazı serimizin ilk yazısında sizlerle birlikte ‘Lineer Regresyon’ konusunun temellerini inceleyeceğiz. Alfabe için harfler, matematik için sayılar, konuşmak için kelimeler ne ifade ediyorsa lineer regresyon (yazının ilerleyen kısımlarında bundan ‘LR’ olarak bahsedilecektir) da makine öğrenimi için o anlama gelmektedir. Şimdi lineer regresyona neden ihtiyaç duyulduğunu temel bir örnek ile anlayalım.



Yukarıdaki grafikte bir düzleme yayılmış 9 farklı noktayı görmektesiniz. Bu grafiğin, bir A ürününün 9 aylık maliyetini (X eksen) ve buna karşılık kaç TL karşılığında satıldığını (Y eksen) temsil ettiğini düşünelim. Örnek olarak yukarıda kırmızı ile işaretlediğimiz noktadaki ürün 1 TL’ye mal edilirken 2 TL’ye satılmış. Sizin amacınız ise, alış satış fiyatları arasındaki ilişkinin yukarıdaki gibi olduğu bir ürün için “Eğer bu ürünün maliyeti 1,5 TL ise satış fiyatı kaç TL olur?” sorusunu cevaplayabilmek. Eğer siz bu noktaları “ $y = ax + b$ ” olarak ifade edilen bir doğru ile temsil edebilirsiniz, o doğru denkleminde x yerine 1,5 yazdığınızda ilgili ürünün kaç TL karşılığında satılacağını da görebilirsiniz. Matematikten de bildiğimiz gibi bir noktadan sonsuz sayıda doğru geçer yani bizim bu örnek problemimiz için şu anda elimizde sonsuz sayıda potansiyel LR doğrusu var. Aşağıdaki şekilde bunlardan birkaç tanesi temsil edilmiştir.



Bu görselde sadece 11 tane örnek doğruyu görüyorsunuz. Sadece bu 9 adet nokta için bile sonsuz sayıda olan doğrulardan rastgele bir tanesini seçip “Aradığımız LR doğrusu budur” diyebilmek biraz zor. Ama yine de konunun pekişmesi için biz bu örnek üzerinden ilerleyelim. Diyelim ki sonsuz sayıda olasılıkları elediniz ve sizin elinizde yalnızca gördüğünüz 11 örnek doğru kaldı. Artık bunlardan bir tanesini LR doğrusu olarak seçip geleceğe yönelik tahminlerinizi yapmanız lazım. Eğer seçiminizde başarılı olduysanız LR doğrunuz, düzlemde bulunan noktaları **hatasız** veya **en düşük hata** ile temsil etmeli. Peki bu hata ile kastettiğimiz nedir? Onu da siyah ile işaretlediğimiz doğruyu inceleyerek öğrenelim.



Yukarıdaki grafikte de net bir şekilde gördüğünüz gibi, tüm noktaların bu siyah doğruya bir uzaklığı var. Bu uzaklıkları kırmızı doğru parçaları ile ifade edelim ve yeşil ile işaretlediğimiz noktayı inceleyelim. Bu noktanın koordinat değerleri (0,3) ancak bizim doğrumuzda $x = 0$ noktasına karşılık gelen y değeri 3 değil 0. Eğer seçilen LR doğrusu örneğimizdeki siyah doğru ise $x = 0$ noktası için 3 TL'lik bir hata var. Veri kümemizde bulunan tüm x noktalarının (gerçek y değeri – doğru üzerindeki y değeri) kadar bir hatası var. Elimizde bulunan tüm hataların toplamı ise LR doğrumuzun hatasını temsil etmekte. Ancak bu hataları toplarken klasik toplama işlemi yapmıyoruz. Bunun nedenini yeşil dikdörtgen içindeki noktaları inceleyerek anlayalım. Gördüğünüz gibi $x = 1$ noktası için elimizde 3 adet y değeri var. Ancak LR doğrumuzda $x = 1$ için tahmin ettiğimiz satış fiyatı yaklaşık olarak 1,2 TL. Bu durumda $x = 1$ noktası için aşağıdaki tablodakine benzer bir hata söz konusu.

Gerçek Y Değeri	Tahminlenen Y Değeri	Hata
1	1,2	-0,2
2	1,2	0,8
3	1,2	1,8

Bu örnekte gördüğünüz gibi hatamızda negatiflik söz konusu. Yani gerçek satış değerinden daha pahalı satılacağı şeklinde bir tahmin söz konusu. Eğer $x = 1$ noktası için hata değerlerini toplarsak 2,4 TL çıkacaktır. Eğer elimizde bulunan tüm noktalar için bu hesabı sürdürürsek hatamız yaklaşık olarak 4 TL çıkacaktır. Aslında hata değeri daha yüksekken negatif değerlerden dolayı beklenenin altında kalacak, bu da sonuçlarımızın doğru olduğu yanılgısına düşmemize neden olacaktır. Bu yüzden hata değerleri toplanırken negatif işaret dezavantajından kurtulmak için “hata kareleri” yani $[(\text{gerçek } y \text{ değeri} - \text{tahminlenen } y \text{ değeri})^2]$ hesabı yapılır. Sonsuz sayıda LR adayı için bu hesap yapıldığında ise en küçük hataya sahip olan doğru LR doğrusu olarak seçilir ve gelecekte yapılacak olan tahminlemelerde kullanılır.

Lineer regresyon ile ilgili temellerin anlatıldığı yazının burada sonuna geliyoruz. Bir sonraki yazımızda makine öğrenimi uygulamalarında kullanılan **scikit-learn** kütüphanesi ile LR uygulaması yapıp hem daha derin matematiksel temellerine değinecek hem de örnek bir uygulama yapacağız. Sonraki yazıda görüşünceye dek kendinize iyi bakın sağlıklı kalın...