

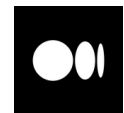
Regresyon Problemleri



istdatascience.com | info@istdsa.com



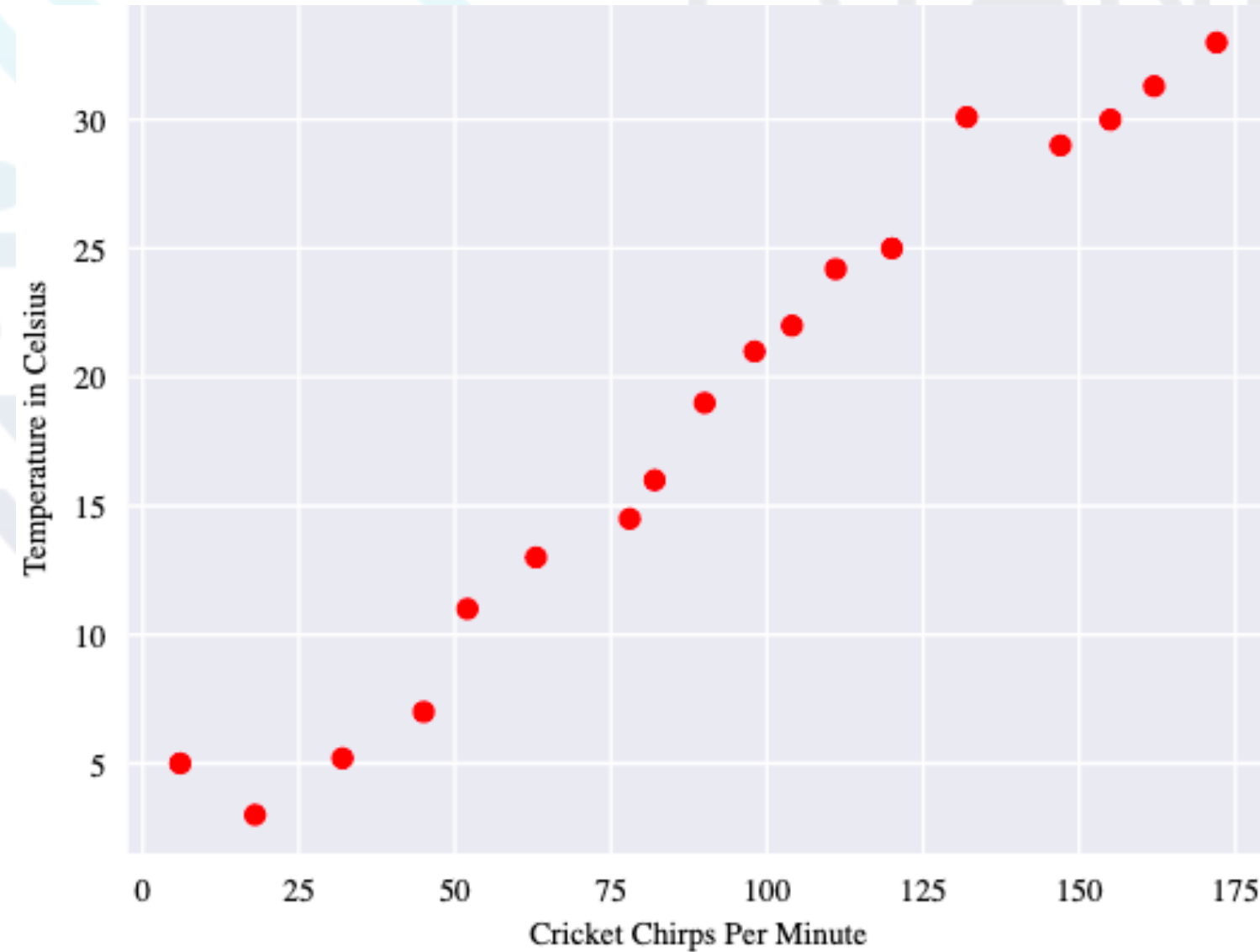
tr.linkedin.com/school/istanbul-data-science-academy



medium.com/istanbuldatascienceacademy

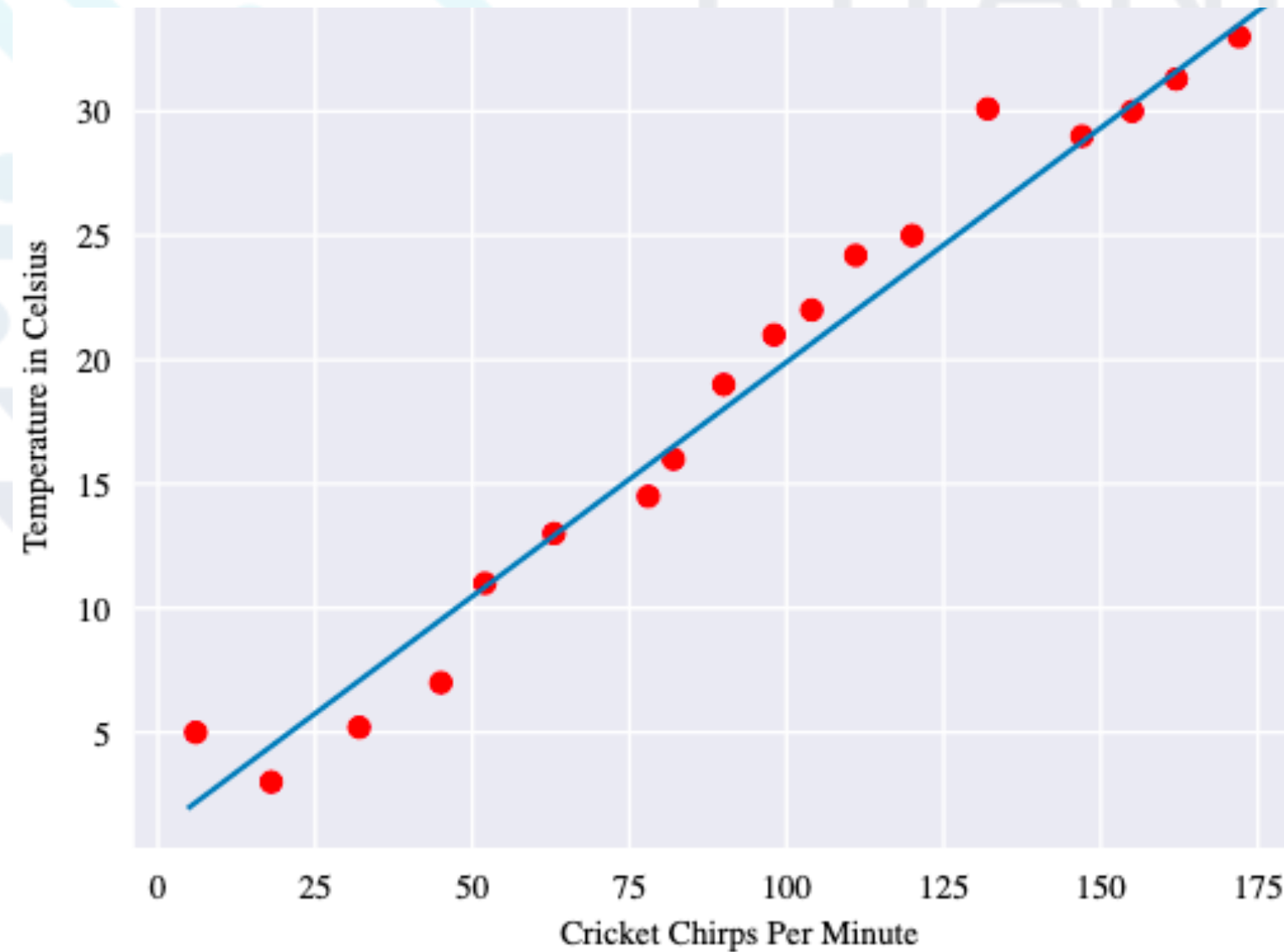
REGRESYON PROBLEMLERİ

Regresyon analizi, verilerdeki eğilimleri bulmanın bir yoludur. Örneğin, kuş cıvıltıları ile sıcaklık değişimi arasında bir bağlantı olup olmadığını kontrol edebilir ve elinizde yeterli bilgi bulunduktan sonra geleceğe yönelik tahminler gerçekleştirebiliriz.



REGRESYON PROBLEMLERİ

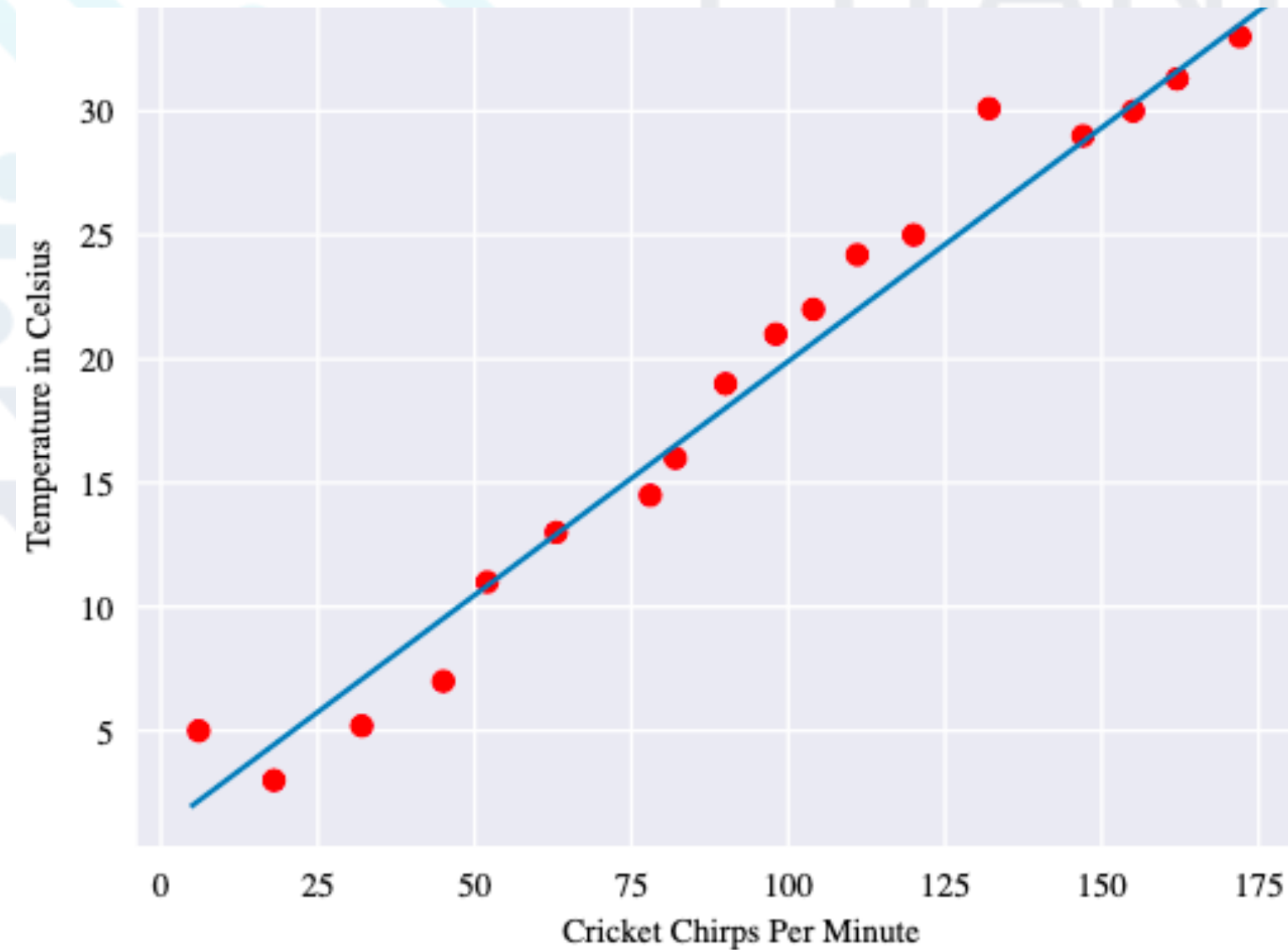
Regresyon analizi, verileriniz hakkında tahminlerde bulunabilmeniz için size bir grafik denklemini sağlayacaktır. İşte bu denklemden regresyon analizinin matematiksel arka planı yatıyor.



REGRESYON PROBLEMLERİ

$$Y = ax + b$$

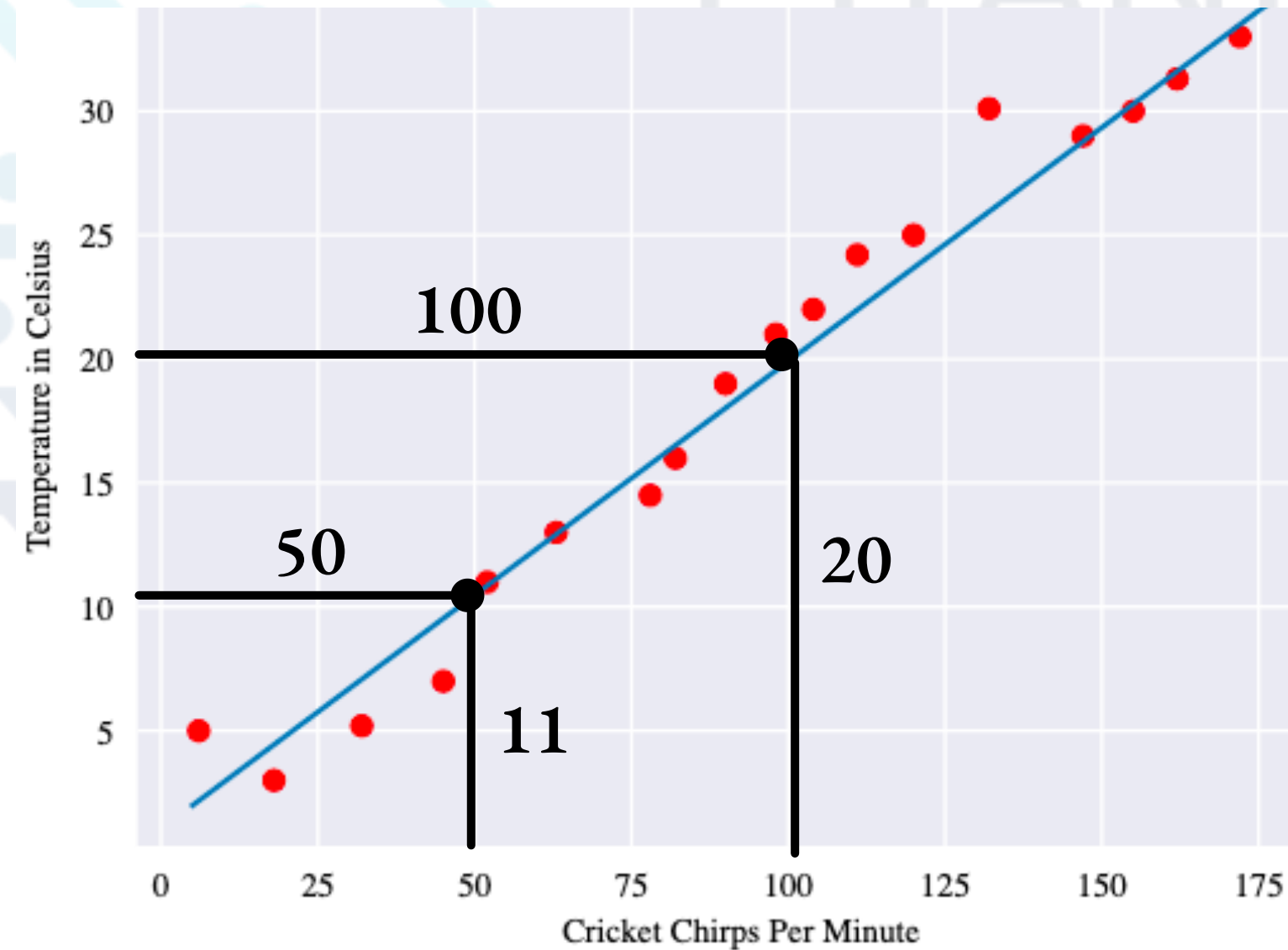
Tahmin Edilen Değer Katsayı Özellik Sabit Değer



REGRESYON PROBLEMLERİ

$$Y = ax + b$$

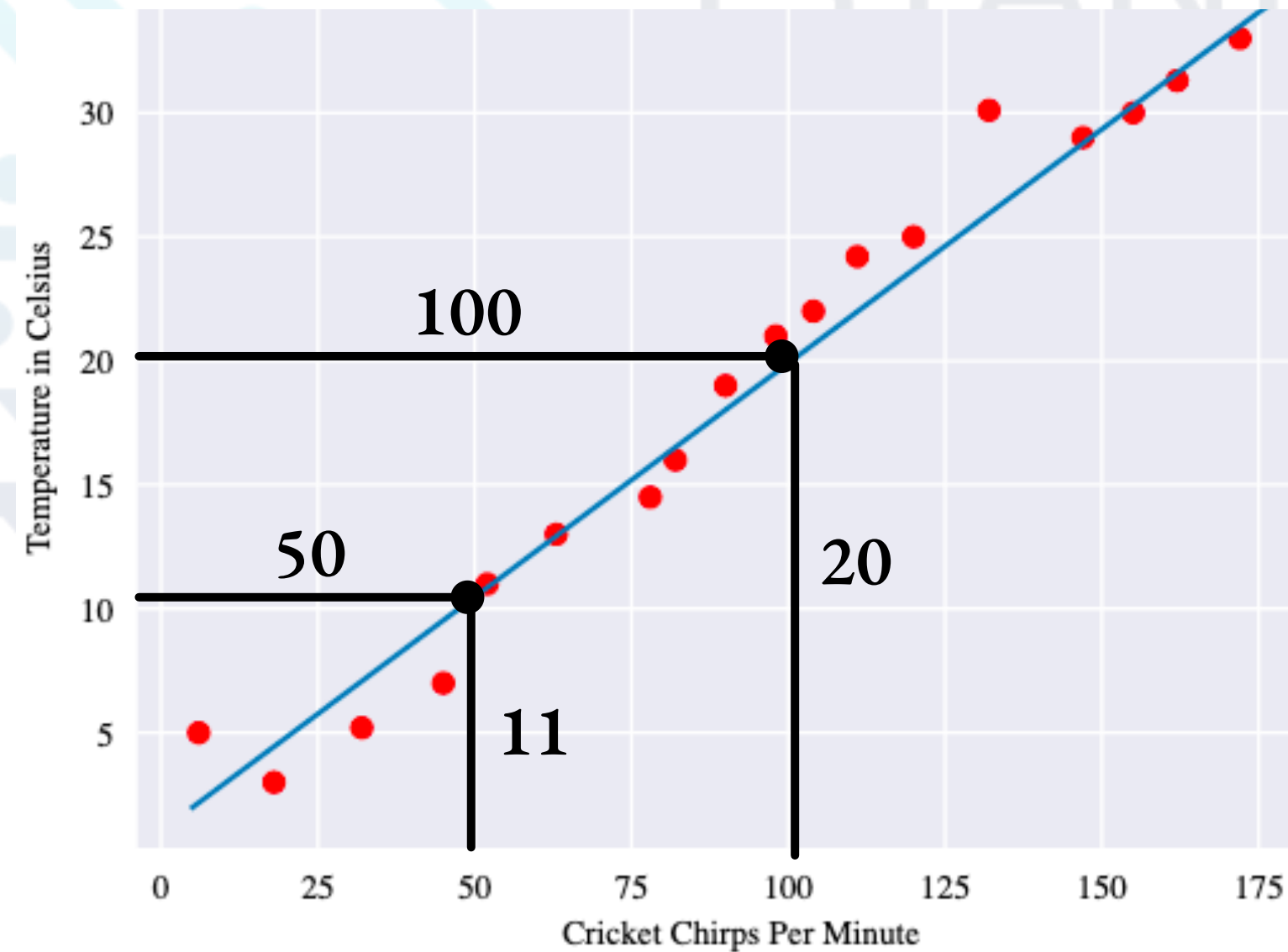
Tahmin Edilen Değer Katsayı Özellik Sabit Değer



REGRESYON PROBLEMLERİ

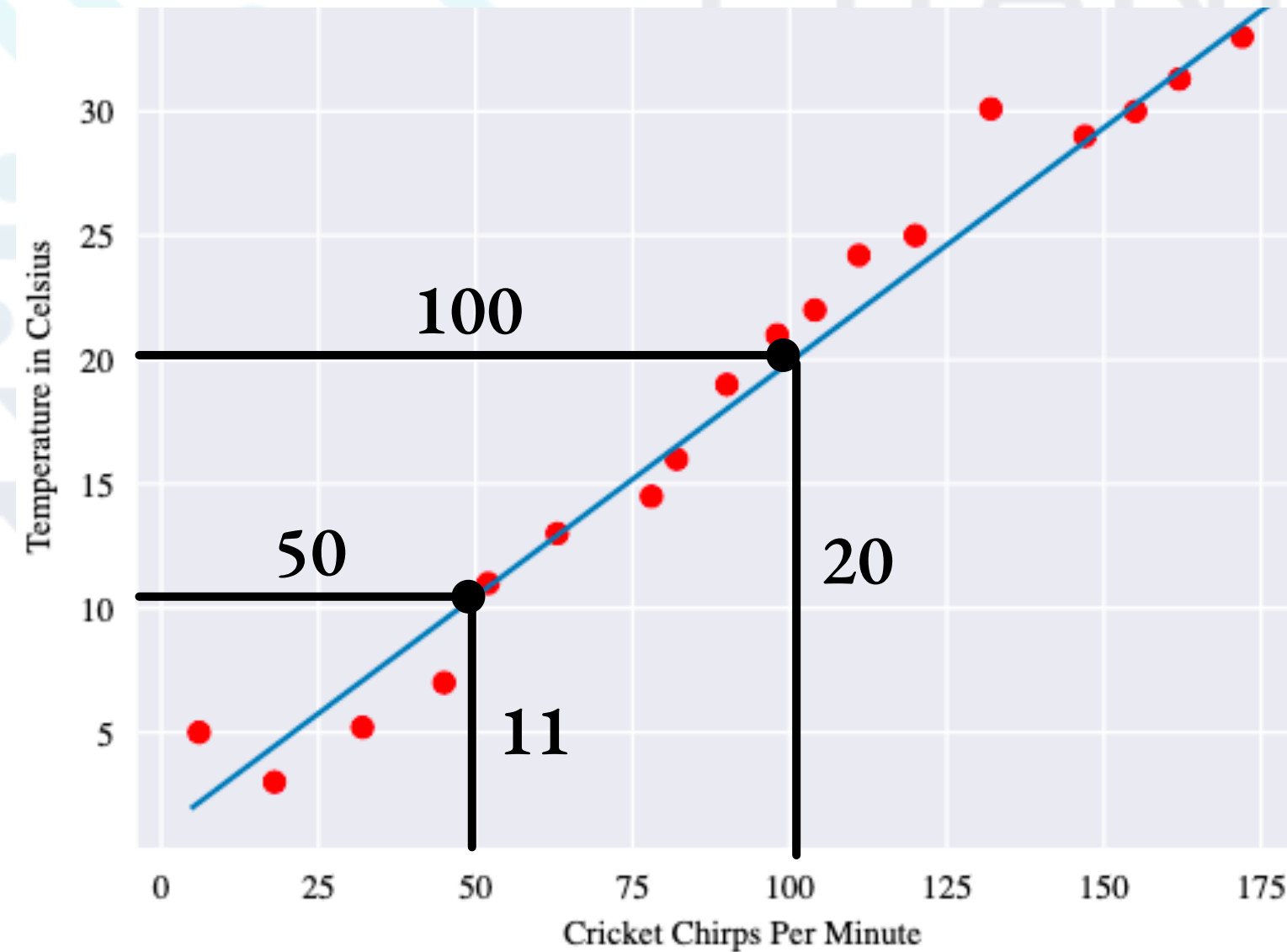
$$20 = a100 + b$$

$$11 = a50 + b$$



REGRESYON PROBLEMLERİ

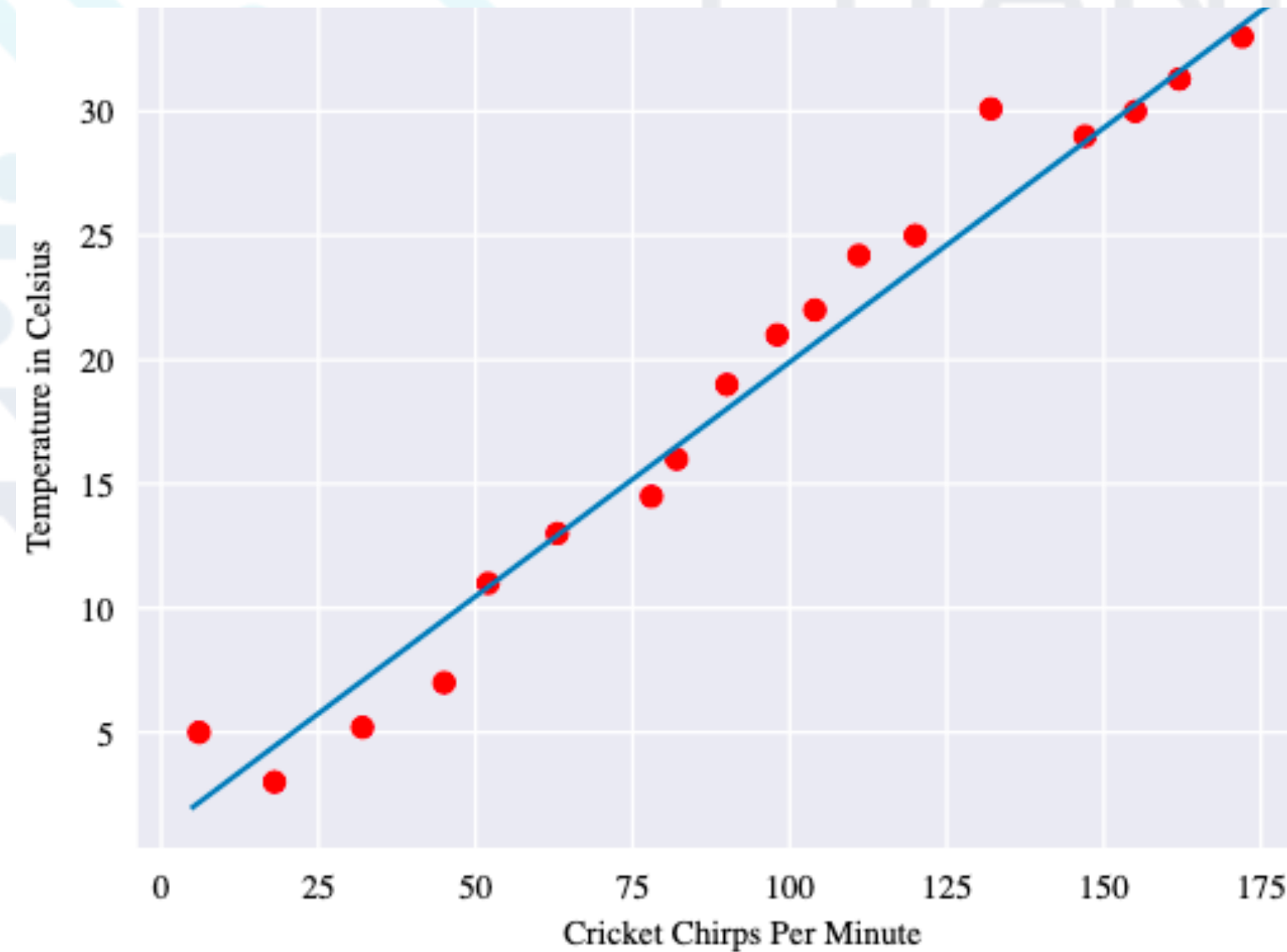
$$\begin{aligned} 20 &= a100 + b \\ 11 &= a50 + b \end{aligned} \rightarrow \begin{aligned} a &= 0.18 \\ b &= 2 \end{aligned}$$



REGRESYON PROBLEMLERİ

$$Y = 0.18x + 2$$

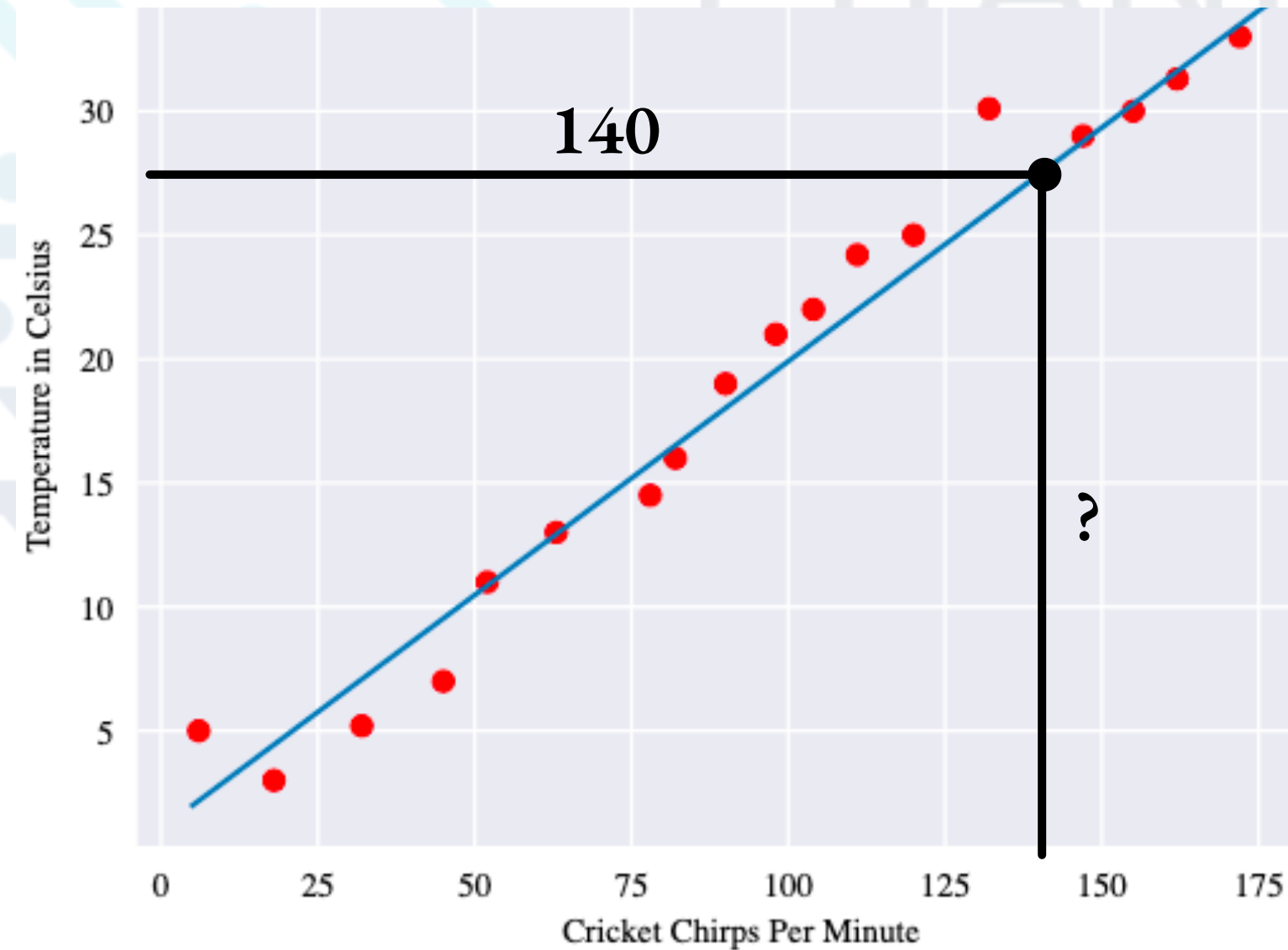
Tahmin Edilen Değer Katsayı Özellik Sabit Değer



REGRESYON PROBLEMLERİ

$$Y = 0.18x + 2$$

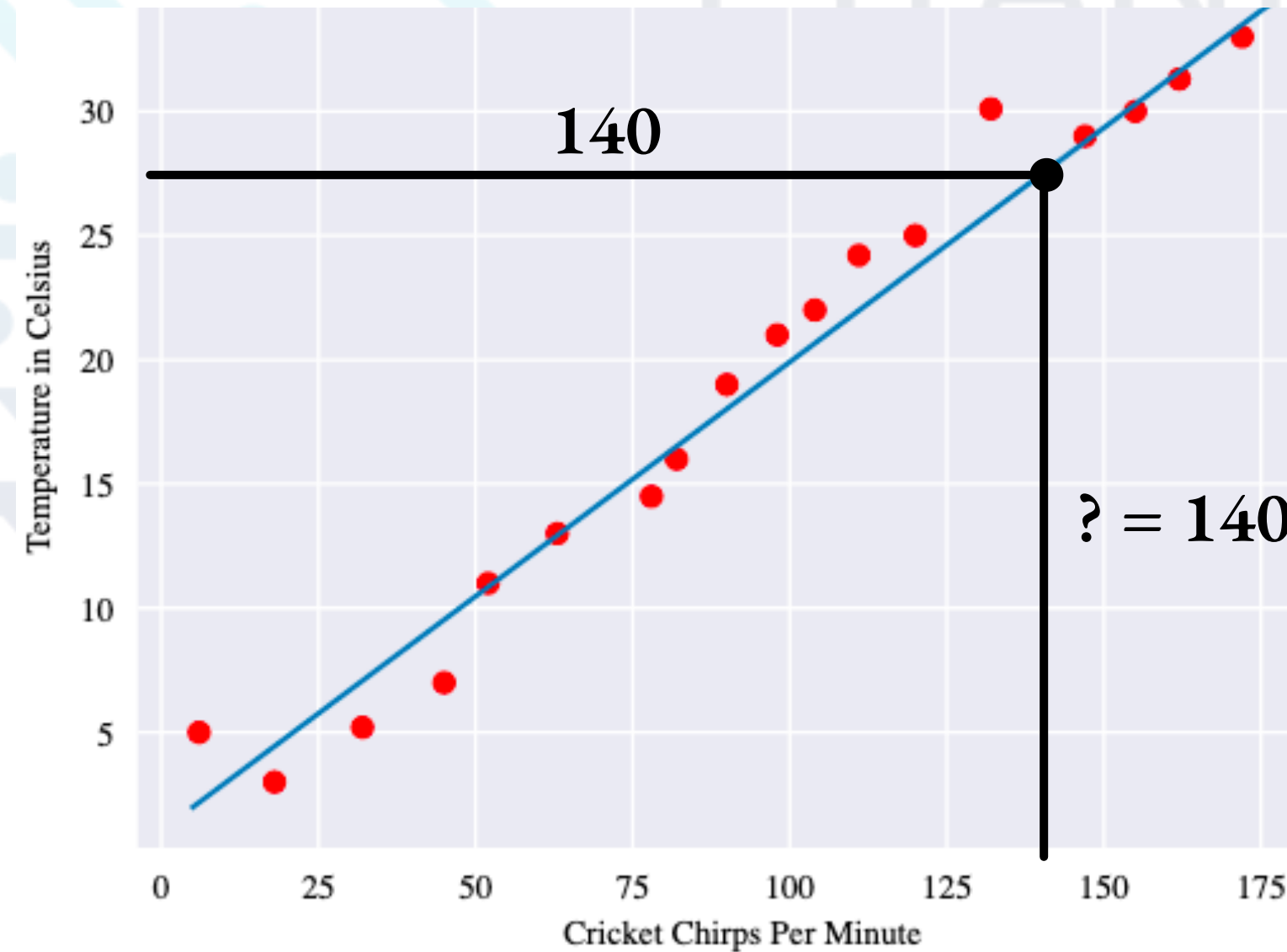
Tahmin Edilen Değer Katsayı Özellik Sabit Değer



REGRESYON PROBLEMLERİ

$$Y = 0.18x + 2$$

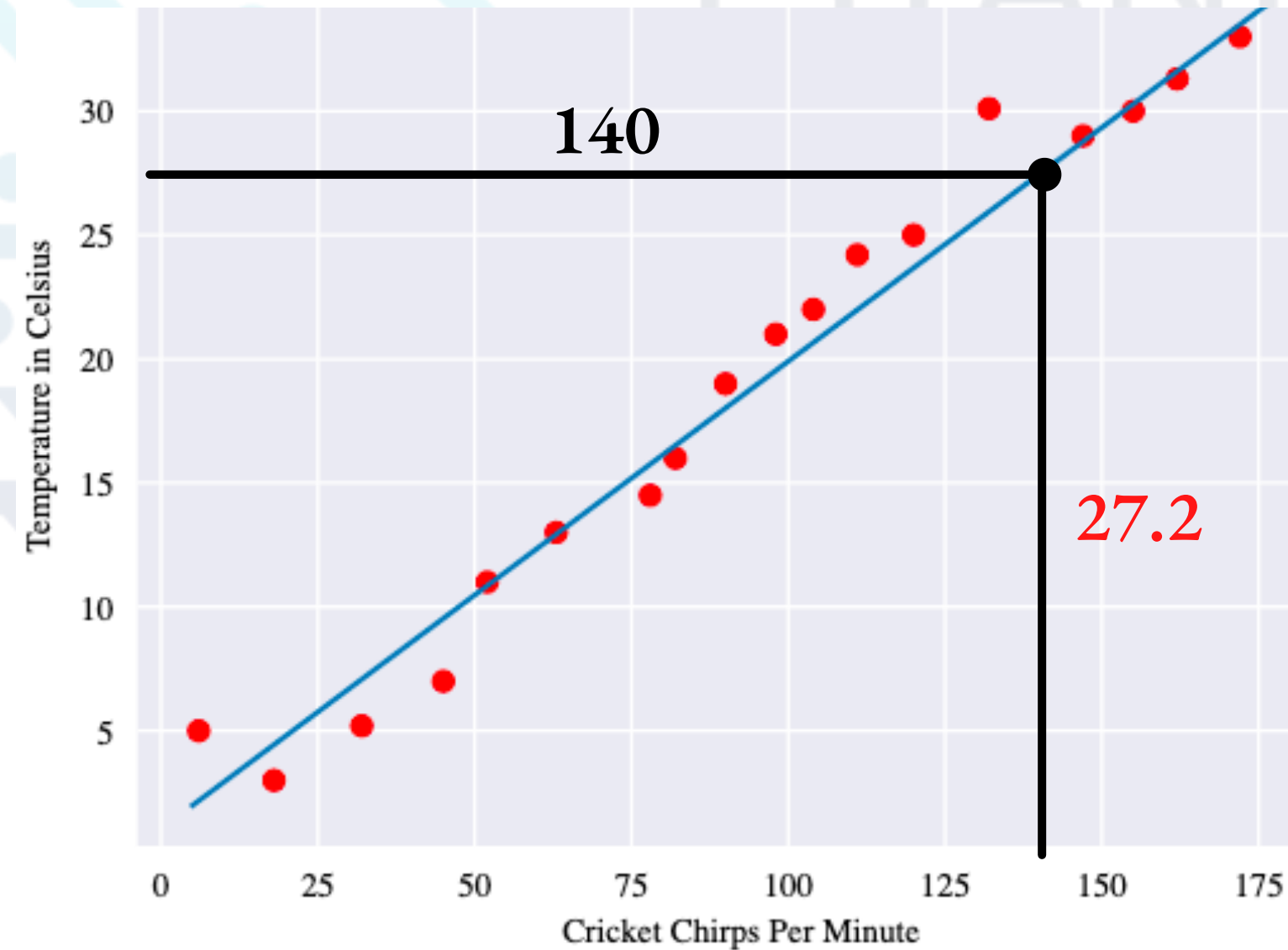
Tahmin Edilen Değer Katsayı Özellik Sabit Değer



REGRESYON PROBLEMLERİ

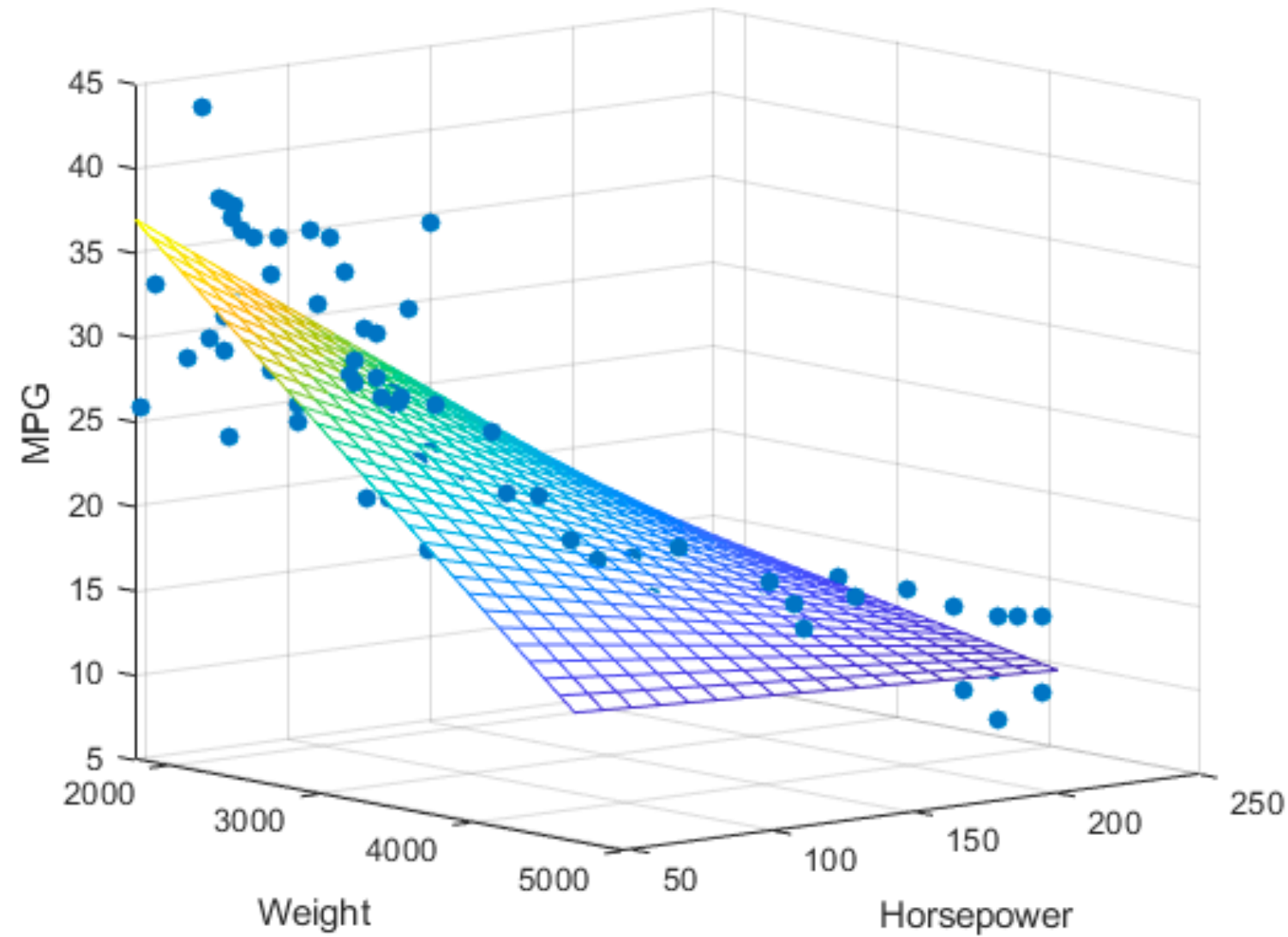
$$Y = 0.18x + 2$$

Tahmin Edilen Değer Katsayı Özellik Sabit Değer



ÇOKLU REGRESYON

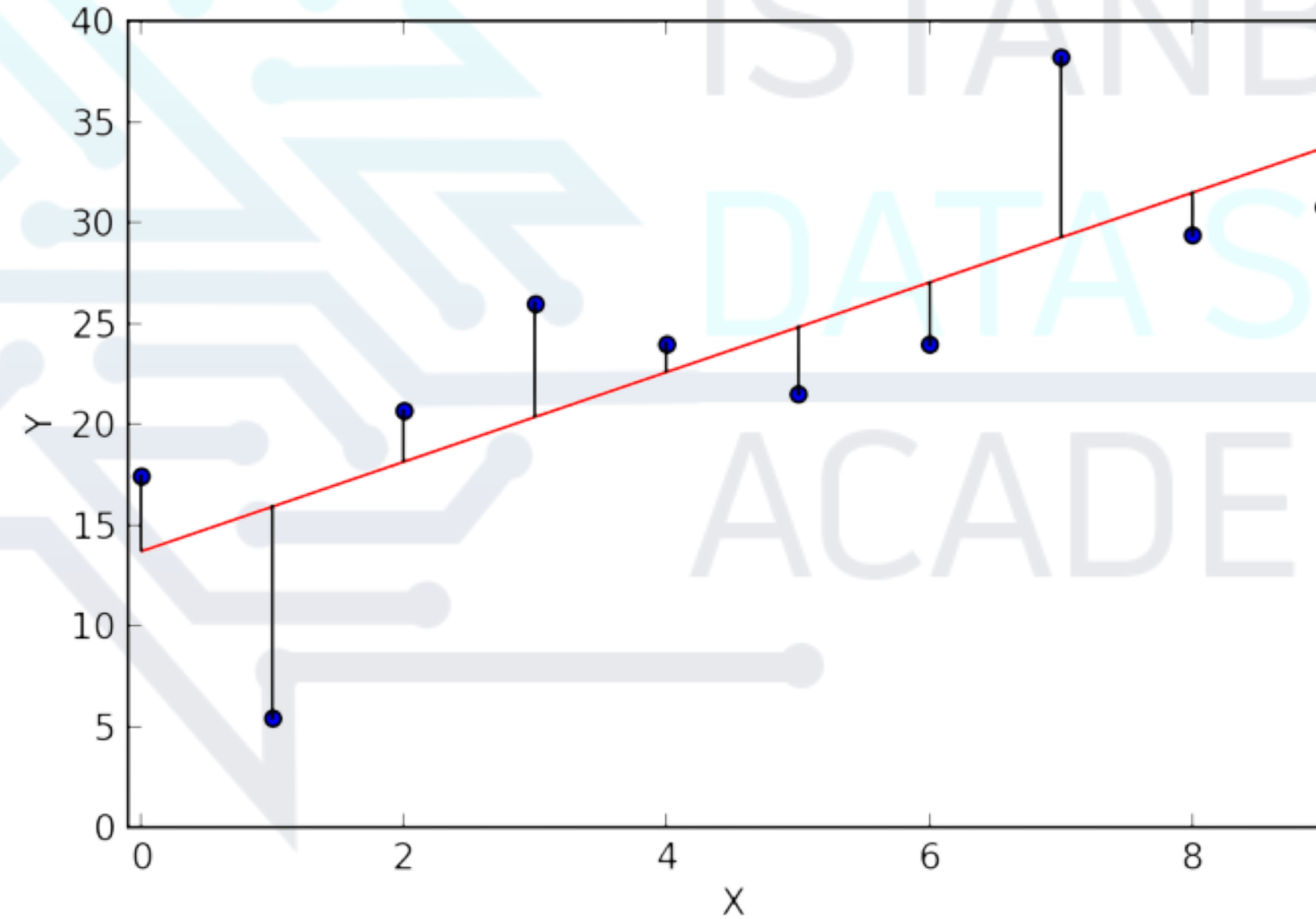
Peki gerçek hayatta sadece tek bir özellik kullanarak ne kadar mantıklı tahminler yapabiliriz? Bu sebeple, genellikle daha karmaşık denklemlerden yararlanarak modeller oluştururuz.



$$Y = ax_1 + bx_2 + cx_3 + \dots + b$$

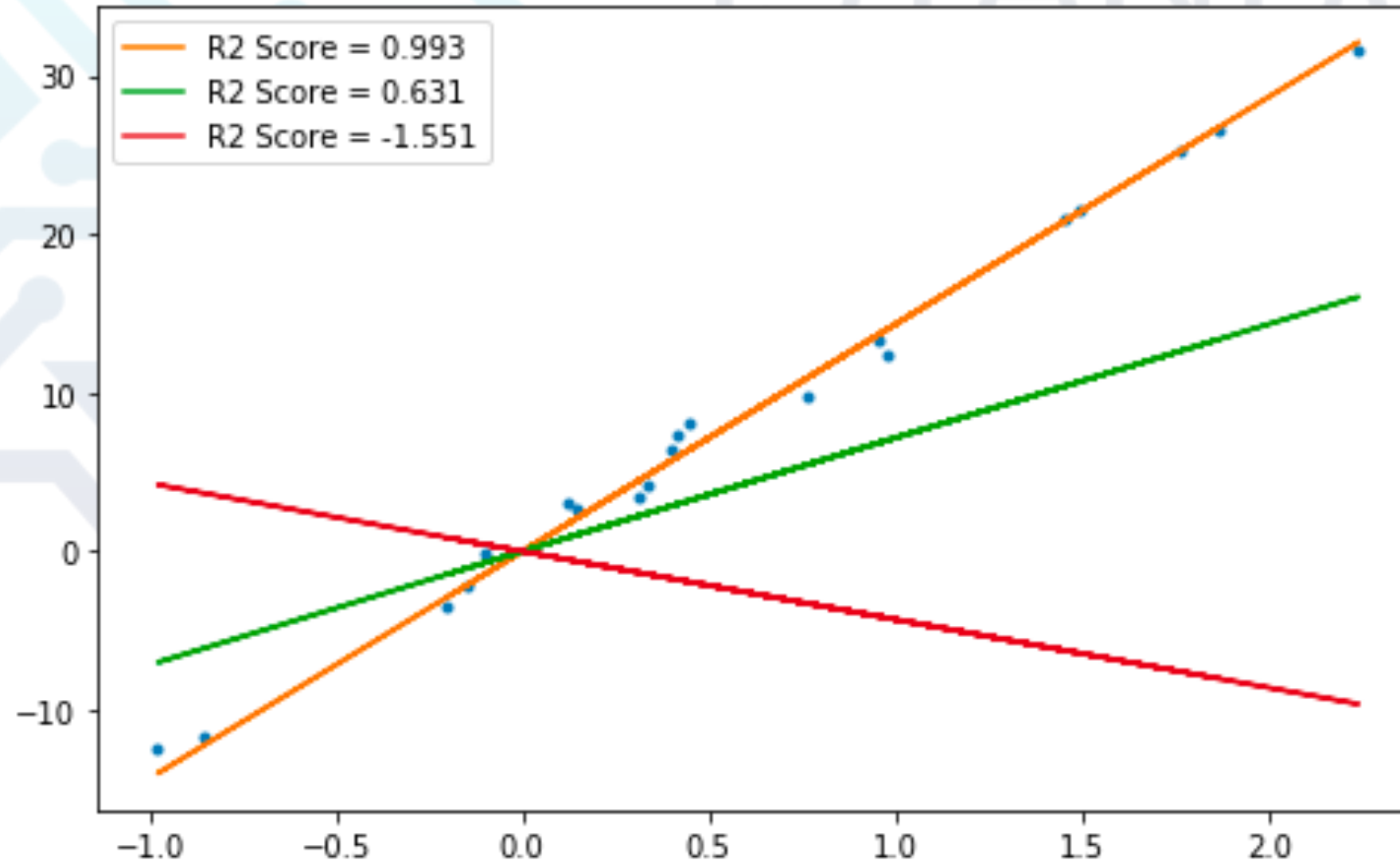
REGRESYON MODELİNİ DEĞERLENDİRME

Regresyon modellerinin doğruluğunu değerlendirmek için modelin yaptığı tahminlerin gerçek verilere ne kadar yakın olduğunu incelememiz gerekir.



R2 SCORE

R2, regresyon modelimiz tarafından belirlenen özelliklerin, hedef değişkenimiz olan Y'ye ilişkin gerçekleştirilen tahminlerin doğruluk yüzdesini temsil eder. Yani olabildiğince 1'e yakın R2 değerleri en çok elde etmek istediğimiz sonuçlardan biridir.



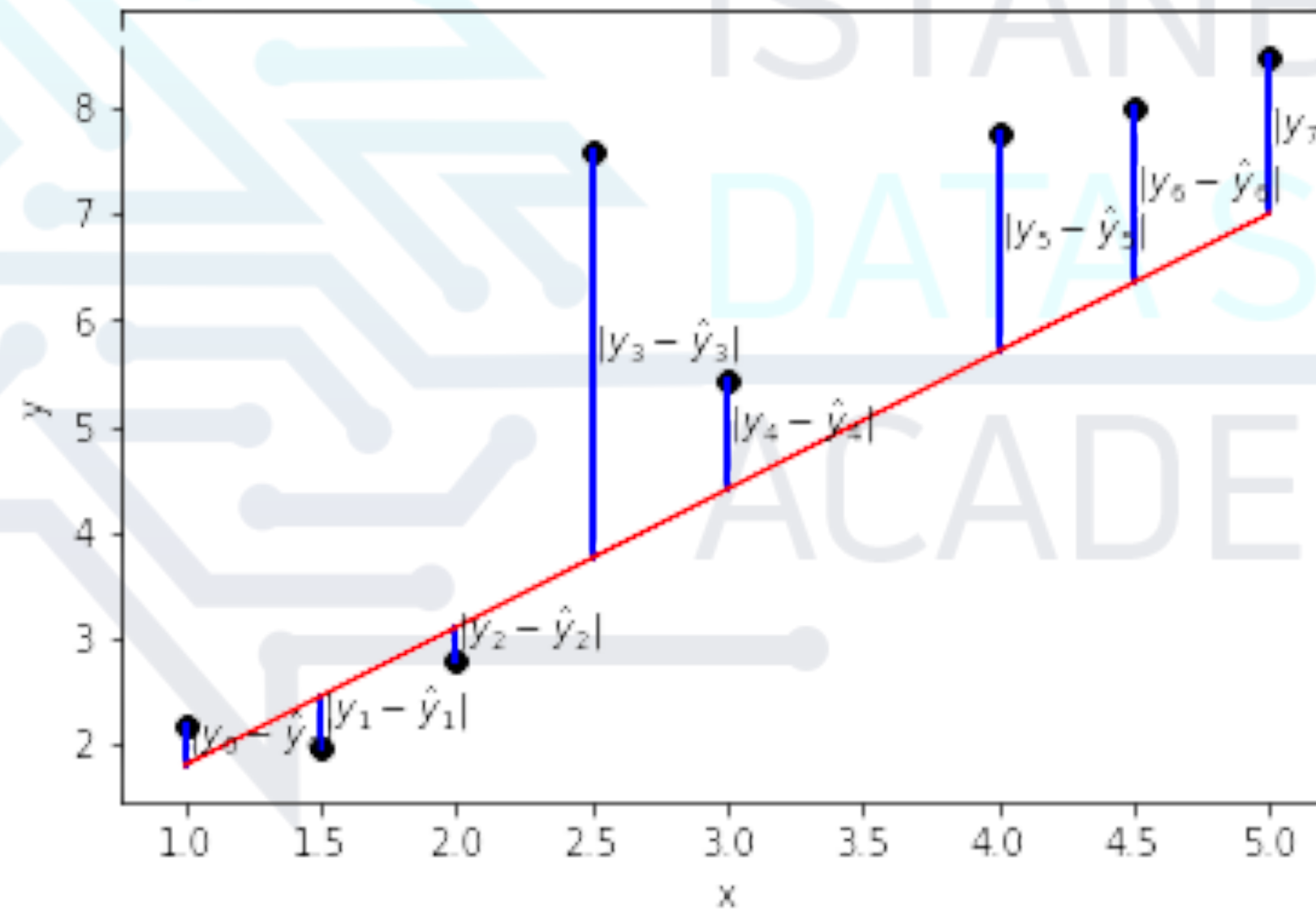
ADJUSTED R2 SCORE

İsmen birbirlerine oldukça benzeseler de Adj. R2 değerinin yorumlanması R2'den biraz daha farklıdır. Adjusted R2 değeri modelimizde bulunan **küçük katsayı değerlerine sahip featureları ve gereksiz featureları** bizim için daha iyi bir model oluşturabilmemiz için **cezalandırır** ve ona göre bir R2 değeri hesabı gerçekleştirir.

```
=====
                        OLS Regression Results
=====
Dep. Variable:          marks      R-squared:          0.971
Model:                  OLS        Adj. R-squared:       0.962
Method:                 Least Squares
Date:                   Sat, 30 May 2020
Time:                   23:44:34   F-statistic:        116.1
No. Observations:      10         Prob (F-statistic):  4.28e-06
Df Residuals:           7         Log-Likelihood:     -17.364
Df Model:               2         AIC:                40.73
Covariance Type:       nonrobust   BIC:                41.64
=====
```

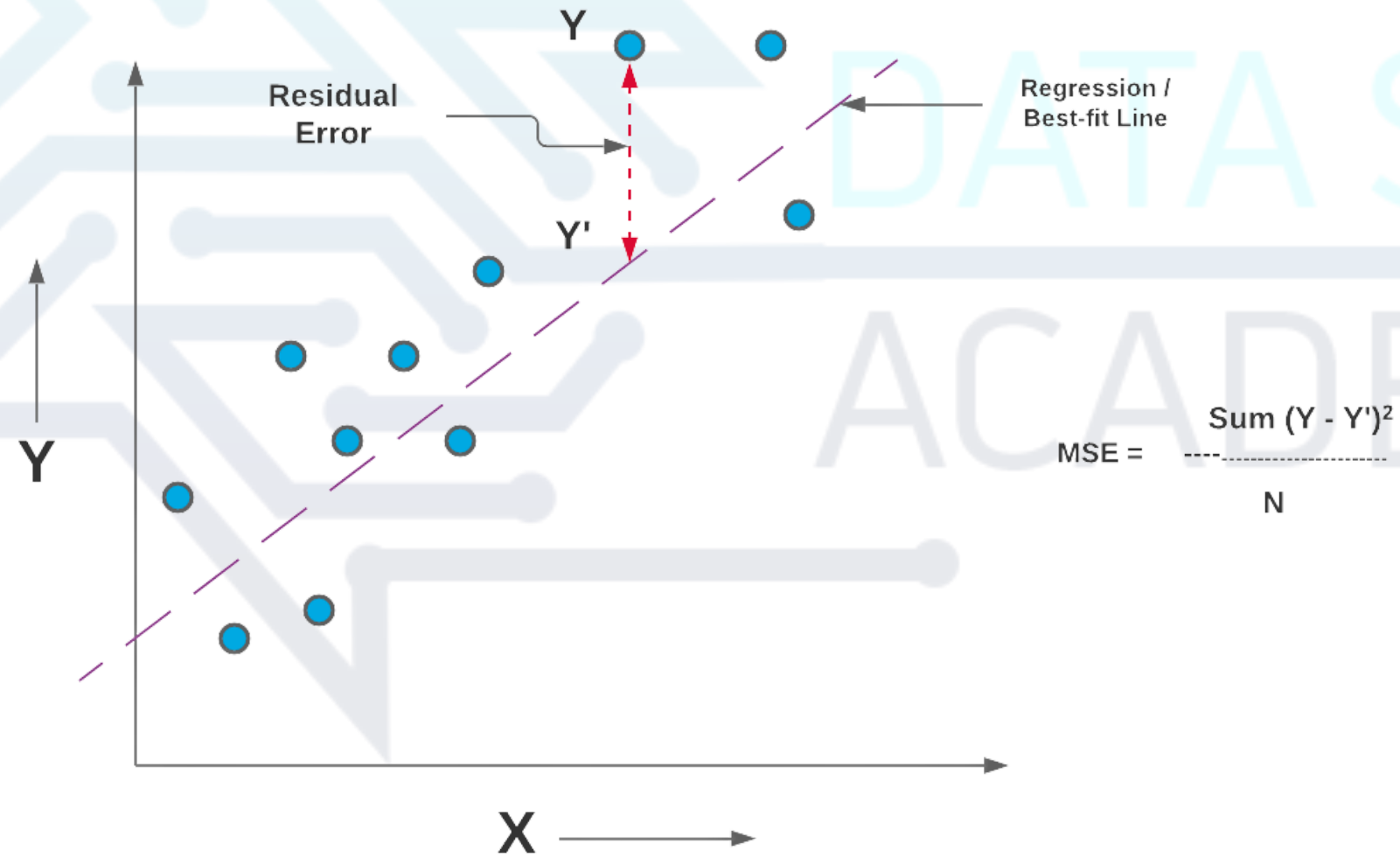
MEAN ABSOLUTE ERROR (MAE)

Mean Absolute Error (MAE), oluşturduğumuz **regresyon modeliyle gerçek verilerimiz arasındaki farkın mutlak değeri** olarak veren hata metriğidir.



MEAN SQUARED ERROR (MSE)

Mean Squared Error (MSE), oluşturduğumuz **regresyon modeliyle gerçek verilerimiz arasındaki mesafenin karesini** alarak elde edilen hata metriğidir. Yüksek miktarda olan hataların tespitinde oldukça başarılıdır.



ROOT MEAN SQUARED ERROR (RMSE)

Root Mean Squared Error (RMSE), oluşturduğumuz regresyon modeliyle gerçek verilerimiz arasındaki mesafenin karesinin karekökünü alarak elde edilen hata metriğidir. Elde edilen hata değerinin yorumlanması daha kolaydır.

