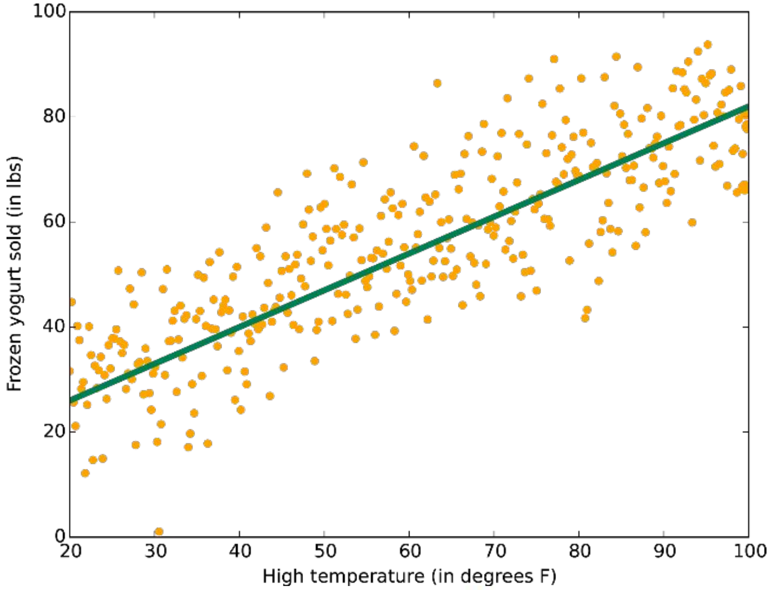
Aprendiz de Machine Learning

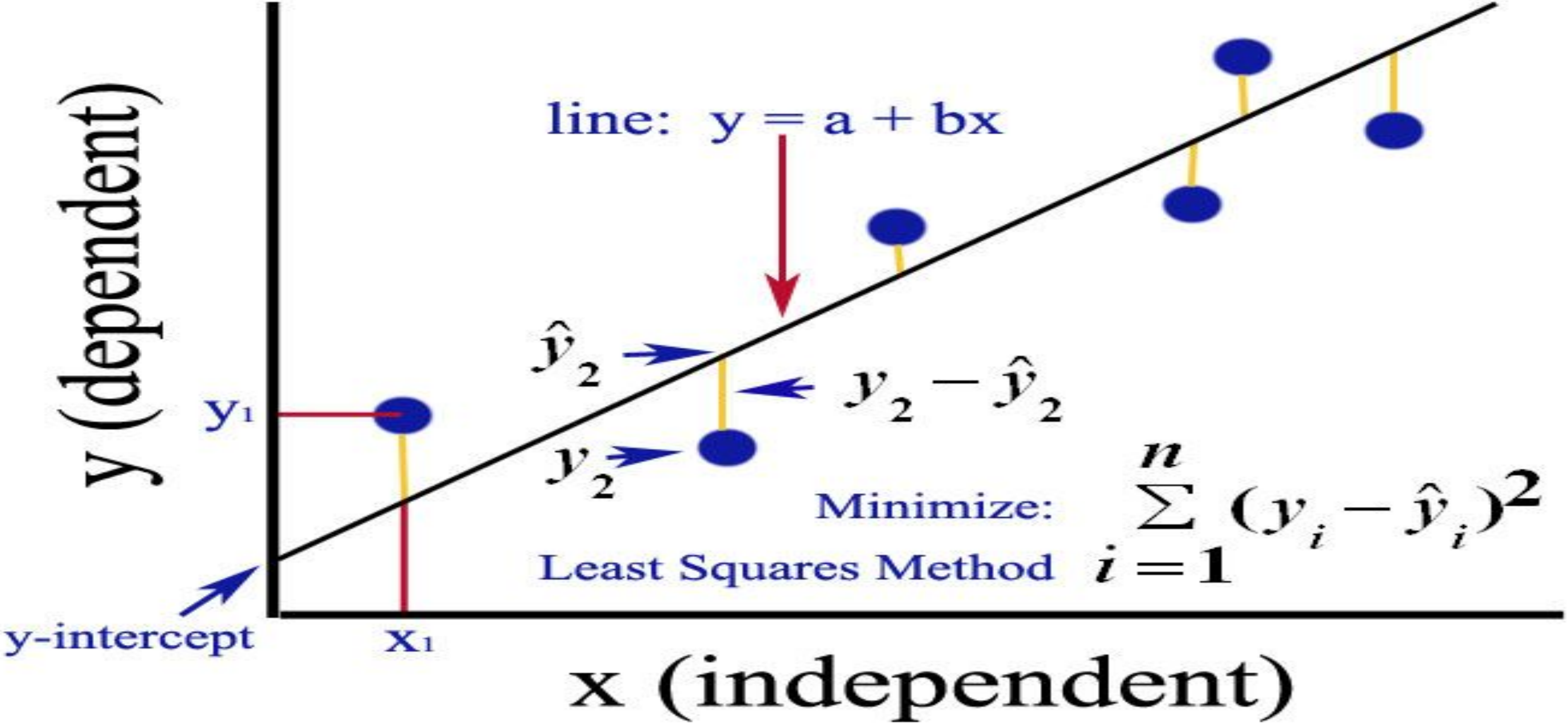
Linear Models for Regression

Linear Models

Linear Models



$$\hat{y} = w^T \mathbf{x} + b = \sum_{i=1}^{p} w_i x_i + b$$


line: y = a + bx
$\hat{y}_2$
$y_2 - \hat{y}_2$
y_2
y_1
Minimize: $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$
Least Squares Method
y-intercept
x_1
y (dependent)
x (independent)

Linear Regression

Ordinary Least Squares

$$\min_{w \in \mathbb{R}^p, b \in \mathbb{R}} \sum_{i=1}^{n} ||w^T \mathbf{x}_i + b - y_i||^2$$

Multicollinearity Problem: when features are correlated, the estimated coefficients becomes highly sensitive to random errors in the observed target.

Ridge Regression

$$\min_{w \in \mathbb{R}^p, b \in \mathbb{R}} \sum_{i=1}^{n} (w^T \mathbf{x}_i + b - y_i)^2 + \alpha ||w||^2$$

The complexity parameter α>=0 impose a penalty on the size of the coefficients. When α is greater the more robust to collinearity problem

Laso Regression

$$\min_{w \in \mathbb{R}^p, b \in \mathbb{R}} \sum_{i=1}^{n} ||w^T \mathbf{x}_i + b - y_i||^2 + \alpha ||w||_1$$

It's useful to reduce the number of features upon which the solution is dependent, because it sets some w coefficients to zero (automatic feature selection)

Scoring

Coefficient of Determination (R^2)

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n-1} (y_i - \bar{y})^2}$$

The coefficient of determination is a statistical measure of how well the regression predictions approximate the real data points. Normally range from 0 to 1, where 1 indicates that the regression predictions perfectly fit the data.

Thank You!