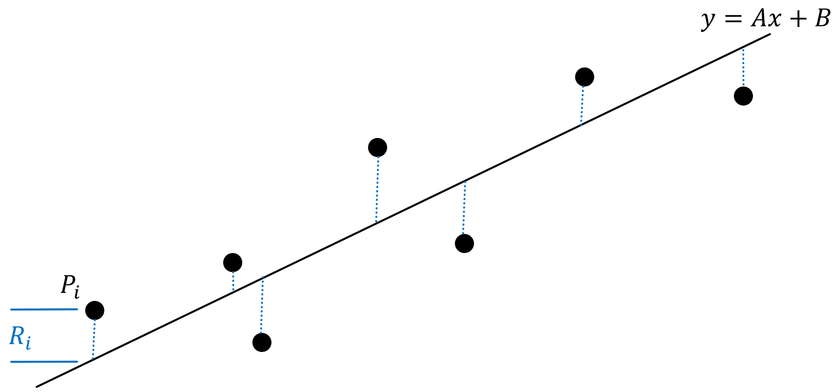
Linear Regression (Slope, Y-Intercept Method)

This method fits a slope, y-intercept line to a series of n >= 2 datapoints.



For this line type, x is the independent variable and y is dependent. Therefore, each value of x must uniquely map to a value of y. This relationship does not allow for the line to be vertical, because a single value of x maps to all the values of y on the line. Additionally, if there are less than 2 unique points the regression becomes singular and cannot be solved.

The slope, y-intercept line equation is:

$$y = Ax + B$$

The general form of least squares regression is to calculate the coefficients to some function F that minimizes the difference between the regressed and measured data.

$$\varepsilon^2 = \sum (F(x) - y')^2$$

Where $F(x) = Ax + B$ and y' are the y-components of the measured values. The equation of least squares linear regression is therefore:

$$\varepsilon^2 = \sum (Ax + B - y')^2$$

The minimum value of ε will be the global minimum of the regression equation. At the global minimum the derivative of ε must be 0.

$$\frac{d\varepsilon}{dy} = 0$$

For the derivative of ε to be 0, the partial derivatives of ε with respect to all the variables in the equation must also be 0. Since the x and y' values are known, we treat them as the coefficients and A and B as the variables. Therefore the partial derivatives of ε with respect to A and B must be 0.

$$\frac{\partial\varepsilon}{\partial A} = 2\sum (Ax + B - y')x = 0$$

$$\frac{\partial\varepsilon}{\partial B} = 2\sum (Ax + B - y') = 0$$

Dividing by 2 on each side of the equation eliminates the 2's outside of the sum. Simplifying the equations and using the distributive law for summations, the equations can be rewritten as:

$$A\sum x^2 + B\sum x = \sum xy'$$

$$A\sum x + B\sum 1 = \sum y'$$

The sum of 1 represents the number of points in the regression (n). Converting the previous equation to matrix form gives.

$$\begin{bmatrix} \sum x^2 & \sum x \\ \sum x & n \end{bmatrix} \cdot \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \sum xy' \\ \sum y' \end{bmatrix}$$

Solve for the coefficients A and B in matrix form.

$$\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \sum x^2 & \sum x \\ \sum x & n \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum xy' \\ \sum y' \end{bmatrix}$$

$$\begin{bmatrix} A \\ B \end{bmatrix} = \frac{1}{n\sum x^2 - (\sum x)^2} \begin{bmatrix} n & -\sum x \\ -\sum x & \sum x^2 \end{bmatrix} \cdot \begin{bmatrix} \sum xy' \\ \sum y' \end{bmatrix}$$

A and B can therefore be calculated algebraically as:

$$A = \frac{n\sum xy' - \sum x \sum y'}{n\sum x^2 - (\sum x)^2}$$

$$B = \frac{-\sum x \sum xy' + \sum x^2 \sum y'}{n\sum x^2 - (\sum x)^2}$$

The linear equation is now:

$$y(x) = Ax + B$$

To measure the accuracy of the fit, we calculate the residuals (R_i) of the equation.

$$R_i = abs(y(x_i) - y_i')$$

The max R_i gives the max deviation of a input point from the regressed equation and the average of all the R_i values gives a good indication of the quality of the fit.