Machine Learning

Tutorships

Nicolas Gartner

Math operators

Final value

The thing to sum

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + a_3 + \cdots + a_{n-1} + a_n$$

Index of summation

First value

Example:

$$\sum_{i=0}^{4} \frac{i}{i+1} = \frac{0}{0+1} + \frac{1}{1+1} + \frac{2}{2+1} + \frac{3}{3+1} + \frac{4}{4+1} = \frac{163}{60} \approx 2.72$$

Mean (arithmetic) or average:

$$A = [a_1, a_2, a_3, \cdots, a_{n-1}, a_n]$$

$$\bar{A} = \mathrm{mean}(A) = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + a_3 + \cdots + a_{n-1} + a_n}{n}$$

Math operators

Measures of the spreading:

Standard deviation:

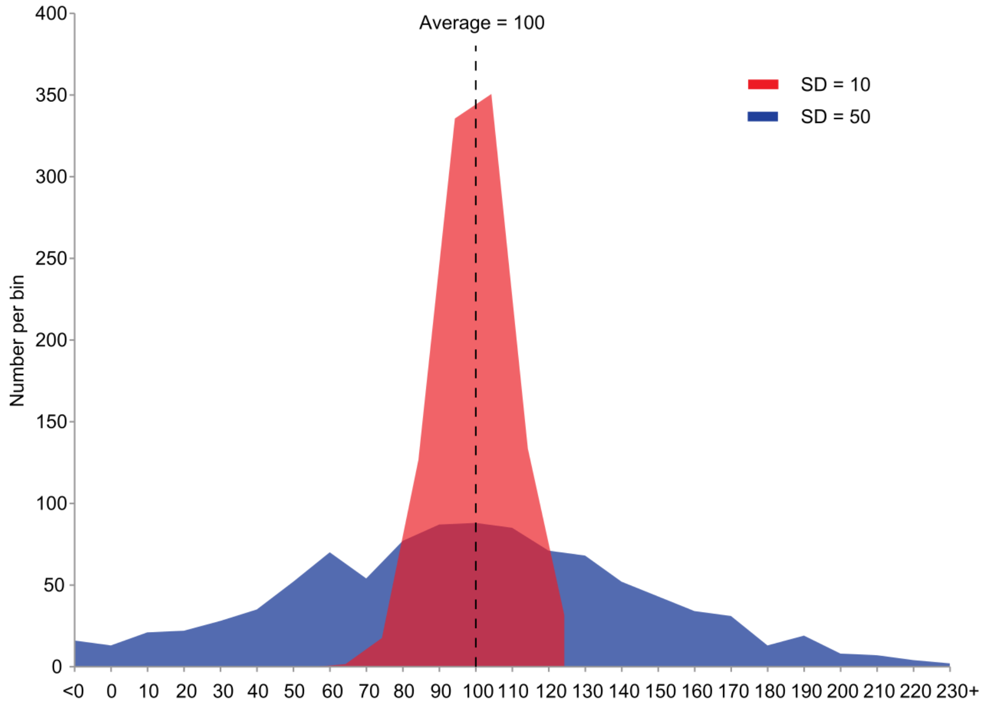
$$A = [a_1, a_2, a_3, \cdots, a_{n-1}, a_n]$$

$$\bar{A} = \text{mean}(A) = \frac{1}{n}\sum_{i=1}^{n} a_i$$

$$\sigma(A) = SD(A) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - \bar{A})^2}$$

Variance:

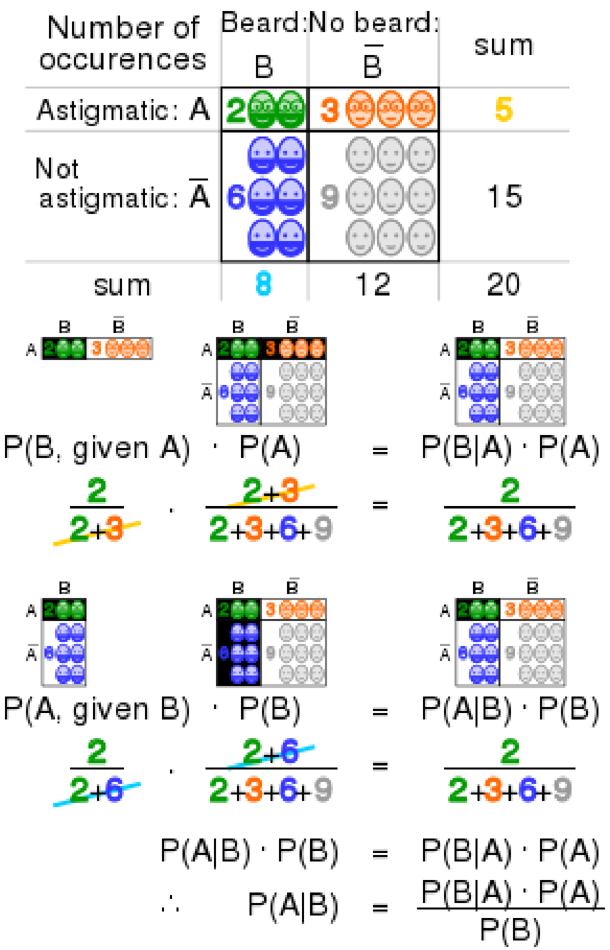
$$\text{Variance}(A) = \left(\sigma(A)\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(a_i - \bar{A})^2$$



If a_i has a unit (m for example):

- $\sigma(A)$ is in m
- but the variance is in m^2

Bayes theorem



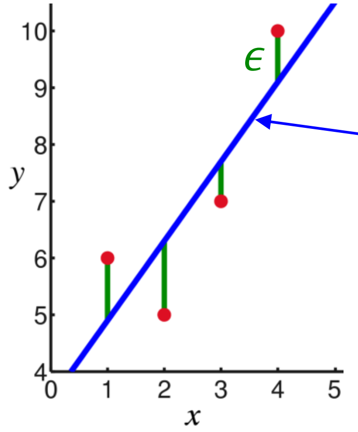
$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

$A \,\&\, B$	are events
$P(A)$	is the probability that A occurs
$P(A \mid B)$	is the likelihood of A occurring given B occurs, called conditional probability

Linear regression principia

Fixed equation from that relates X (input) to y (output) with the acceptance of ϵ (an irreducible error)

$$y = f(X) + \epsilon$$



The function that relates x to y could be:

- $f(X) = \alpha X + \beta$
- $f(X) = \alpha X + \theta X^2 + \beta$
- $f(X) = \alpha e^{-i\omega X} + \beta$

For the rest of the presentation, we will only focus on that simple case

Linear regression objective is to estimate the best matching $f(X)$ which is called estimate, written $\hat{y} = \widehat{f(X)}$

So the objective is to minimize the residual error $e = y - \hat{y}$ for every sample X

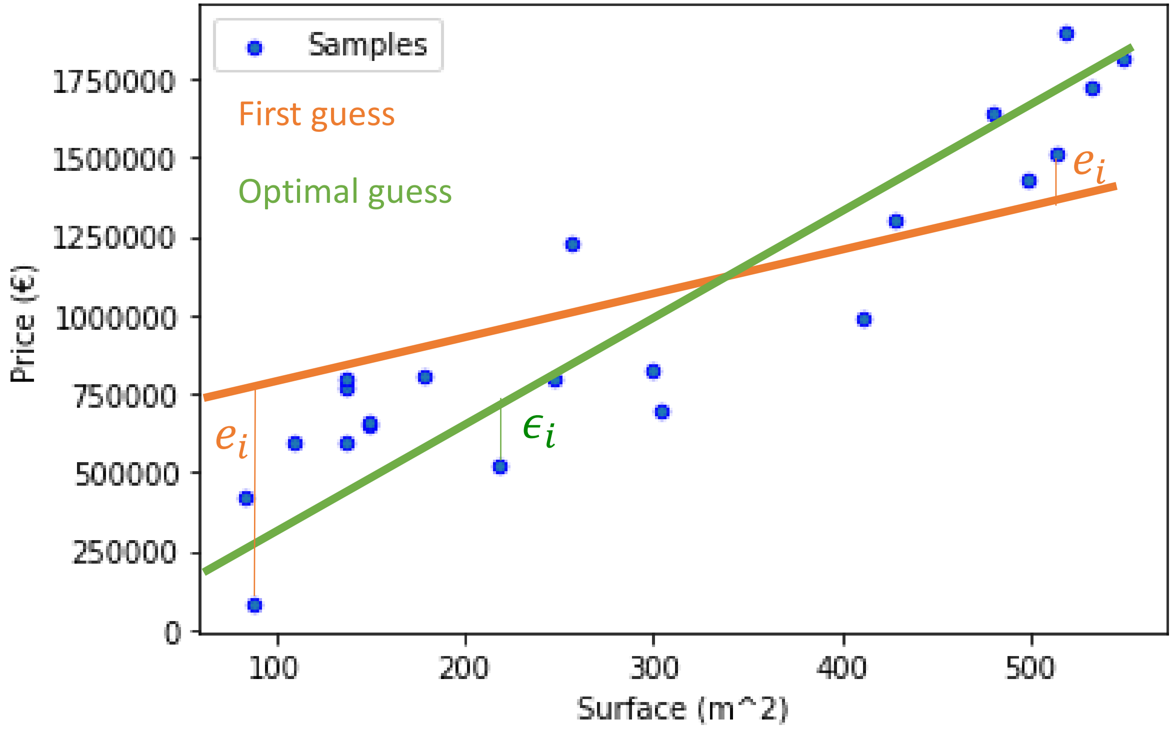
Don't confuse e with ϵ:

$$e = (\alpha X + \beta) + \epsilon - (\hat{\alpha} X + \hat{\beta}) = (\alpha - \hat{\alpha})X + (\beta - \hat{\beta}) + \epsilon$$

Don't confuse e with ϵ

Irreducible error

$$e = (\alpha X + \beta) + \epsilon - (\hat{\alpha} X + \hat{\beta}) = (\alpha - \hat{\alpha})X + (\beta - \hat{\beta}) + \epsilon$$



$$y_i = \hat{\alpha} x_i + \hat{\beta} + e_i$$
$$e_i = y_i - (\hat{\alpha} x_i + \hat{\beta})$$

$$minimize \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

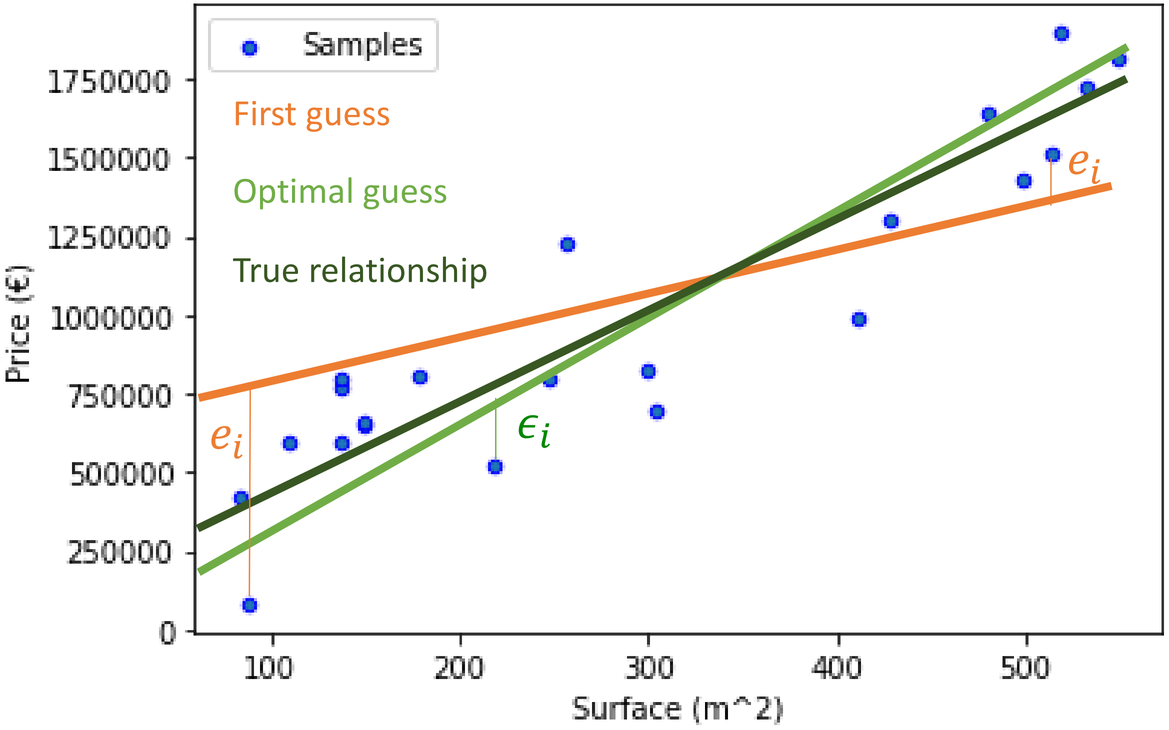
- Using e_i^2 allows to get only positive values for the error
- We do not minimize e_i^2 but we minimize D the sum of all e_i^2

$$D = \sum_{i=1}^{n} e_i^2$$

Optimal guess is obtained when all $e_i \approx \epsilon_i$

The true/initial relationship

The true relationship behind is not always the one that will be obtained with the regression algorithm



$$y_i = \hat{\alpha} x_i + \hat{\beta} + \varepsilon_i$$

α_{true} not always$=\alpha$ β_{true} not always$=\beta$

True relationship can be different than optimal guess:

- The data spreading around the true function is not even
- The more data you have doesn't mean the closer to the true relationship you are (only depends on evenness)

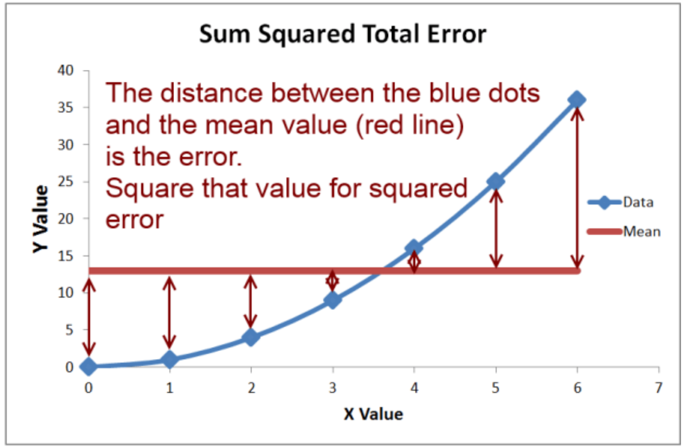
R^2 the coefficient of determination

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

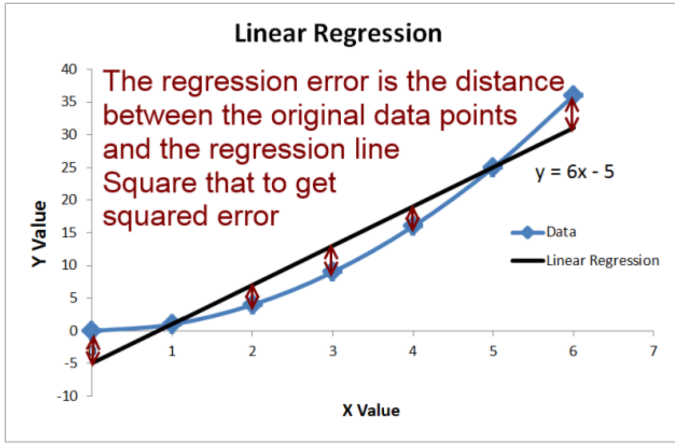
y_i is the true/known value

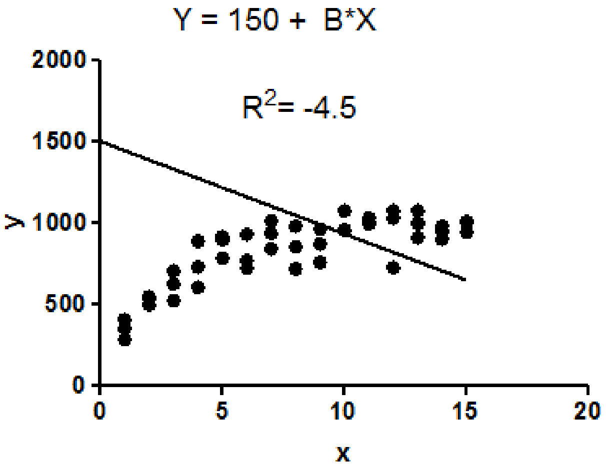
- R^2 compares the fit of the chosen model with that of a horizontal straight line placed at $\bar{y}$ (the mean of Y).
- Negative R^2: the chosen model fits worse than a horizontal line

Sum Squared Total(SST)= $\sum_{i=1}^{n}(y_i - \bar{y})^2$



Sum Squared Regression(SSR)= $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$





$$\boxed{R^2(y, \hat{y}) = 1 - \frac{SSR}{SST}}$$

Perfect regression means SSR=0