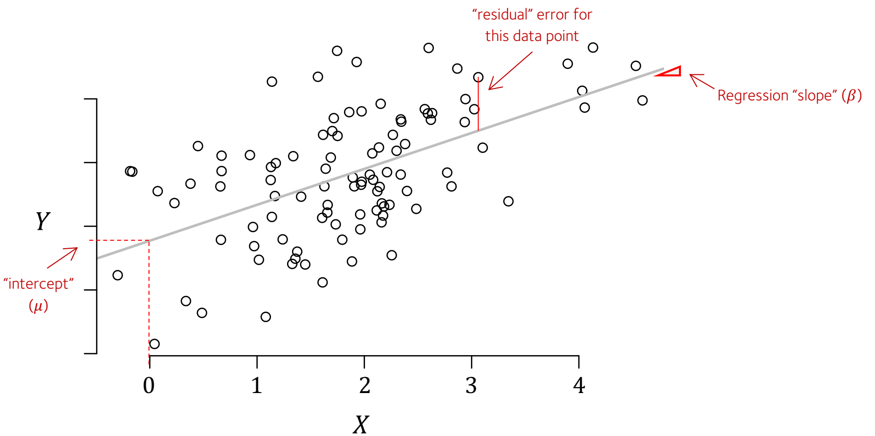
Linear regression models an outcome variable (Y) in terms of one or more predictor variables (X). The model asserts that Y is a linear combination of columns of X plus some noise. The noise is assumed to be Gaussian with some variance σ^2 (that is the same for all data points).

$$Y = \mu + X\beta + \epsilon \qquad \epsilon \sim N(0, \sigma^2)$$

Matrix multiplication if X is multidimensional

(i.e. $Y = \mu + X_1\beta_1 + X_2\beta_2 + \cdots X\beta + \epsilon$)



The likelihood function. An equivalent way to write the above formula is to specify the likelihood function. It is simplest to subsume the parameter μ into the parameters β by assuming the first column of X is the constant column equal to 1:

Likelihood: $$P(Y|X,\beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\frac{\sum_n (Y_n - X_n\beta)^2}{\sigma^2}}$$ Sum of squared errors

Or (matrix notation): $$P(Y|X,\beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\frac{(Y-X\beta)^t(Y-X\beta)}{\sigma^2}}$$