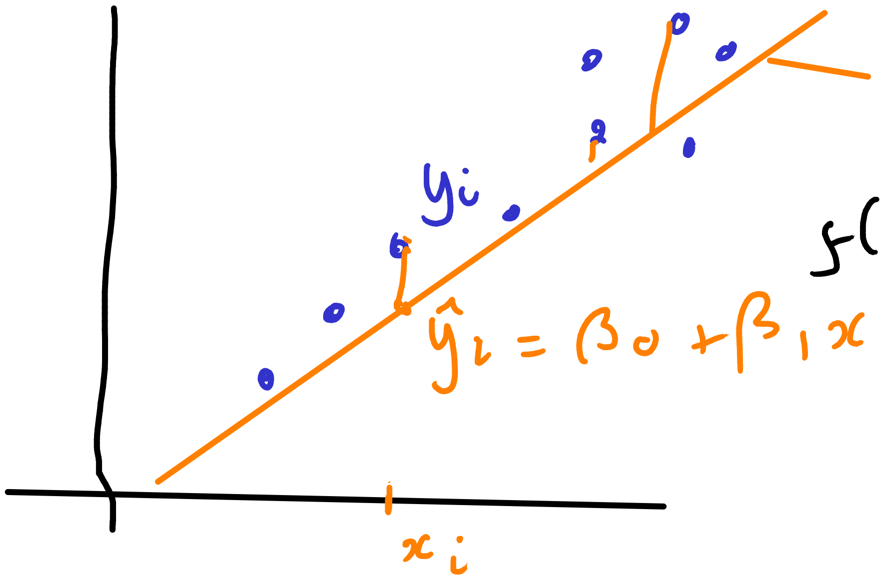
Foundations of Data Science: Logistic regression - Maximum likelihood estimation of logistic regression coefficients

Principle of Maximum Likelihood



$$f(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 + \beta_1 x)^2$$

Optimisation:

$\beta_0 =$ $\qquad\qquad$ $\beta_1 =$

Adjust coefficients so as to maximise the likelihood of the data.

$\Rightarrow$ Expression for max. likelihood

Optimise w.r.t $\beta_0, \beta_1, \ldots$

Maximum likelihood function

$y_i = 1 \Rightarrow$ Success $\qquad y_i = -1 \Rightarrow$ Failure

$$P(Y = y_i \mid X = x_i) = \frac{y_i + 1}{2} f(\beta_0 + \beta_1 x_i) + \frac{-y_i + 1}{2}\left(1 - f(\beta_0 + \beta_1 x_i)\right)$$

$$\left[1 - f(\beta_0 + \beta_1 x_i) = f(-\beta_0 - \beta_1 x_i)\right]$$

$$P(Y = y_i \mid X = x_i) = \frac{y_i + 1}{2} f(\beta_0 + \beta_1 x_i) + \left(\frac{-y_i + 1}{2}\right) f(-\beta_0 - \beta_1 x_i)$$

$$= \frac{y_i + 1}{2} f(y_i(\beta_0 + \beta_1 x_i)) + \left(\frac{-y_i + 1}{2}\right) f(y_i(\beta_0 + \beta_1 x_i))$$

$$= f(y_i(\beta_0 + \beta_1 x))$$

Likelihood of dataset given model

$$P(Y = y \mid X = \underline{x}) = P(Y = y_1 \mid X = x_1)\, P(Y = y_2 \mid X = x_2) \ldots$$

$$= \prod_{i=1}^{n} P(Y = y_i \mid X = x_i)$$

$$= \prod_{i=1}^{n} f(y_i (\beta_0 + \beta_1 x_i))$$

$$\ln ab = \ln a + \ln b$$

$$\ln \prod_{i=1}^{n} a_i = \sum_{i=1}^{n} \ln a_i$$

Log likelihood

$$\ln P(Y = y \mid X = \underline{x}) = \sum_{i=1}^{n} \ln f(y_i (\beta_0 + \beta_1 x_i))$$

Optimisation $\Rightarrow$ $\hat{\beta}_0$ and $\hat{\beta}_1$.

https://covid.joinzoe.com/