Logistic Regression

$$\frac{d}{dt} W(t) = \beta W(t) \quad = \quad W(t) = A e^{\beta t}$$

-1874   -1849

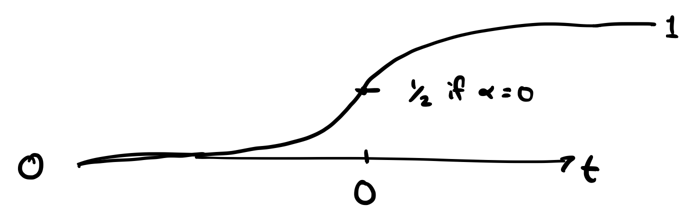
Quetelet & Verhurst

population limit

$$\frac{d}{dt} W(t) = \beta W(t) (\Omega - W(t))$$

let $P(t) = \dfrac{W(t)}{\Omega}$

$$\frac{d}{dt} P(t) = \beta P(t)(1 - P(t)) \Rightarrow P(t) = \frac{e^{(\alpha + \beta t)}}{1 + e^{(\alpha + \beta t)}}$$



"logistic function"

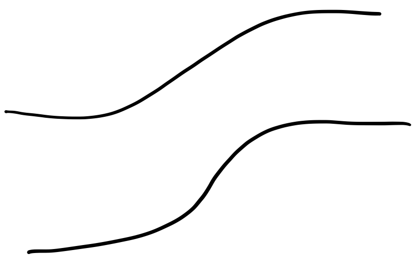
9.5 ~~6.6~~ million Belgium → 11 million

40 million France → 66 million

1920 Pearl + Reed

Ω for US 197 million $\rightarrow$ 312

Yule



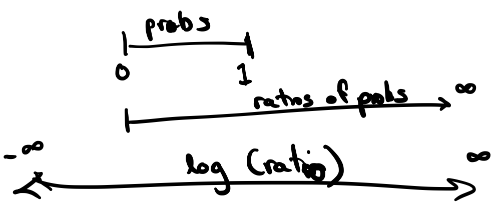
Joseph Berskon 1899 - 1982

$x \in \mathbb{R}^p$ deterministic, not random

$Y \sim \text{Bernoulli}(P(Y=1 \mid x))$ $\quad Y = \pm 1 \rightsquigarrow \overset{\text{0 or -1}}{\tilde{0}}$ or 1

linear model $\lambda^T x$

$$\ln\Bigg(\underbrace{\frac{P(Y=1 \mid x, \lambda)}{P(Y=\tilde{0} \mid x, \lambda)}}_{\text{odds ratio}}\Bigg) = \lambda^T x \longrightarrow \frac{P(Y=1 \mid x, \lambda)}{P(Y=\tilde{0} \mid x, \lambda)} = e^{\lambda^T x}$$



$$P(Y=1 \mid x, \lambda) = e^{\lambda^T x}\left(1 - P(Y=1 \mid x, \lambda)\right)$$

$$P(Y=1 \mid x, \lambda)\left[1 + e^{\lambda x}\right] = e^{\lambda x}$$

$$P(Y=1 \mid x, \lambda) = \frac{e^{\lambda x}}{1 + e^{\lambda x}}$$

logistic function

$$L(\lambda) = P(Y_1 = y_1, \dots, Y_m = y_n \mid \lambda, x_1 \dots x_n)$$

$$= \prod_{i=1}^{n} P(Y_i = y_i \mid \lambda, x_i)$$

$$\lambda^* \in \underset{\lambda}{\operatorname{argmax}}\ L(\lambda) = \underset{\lambda}{\operatorname{argmax}}\ \log L(\lambda)$$

Simplify:

$$P(Y = y_i \mid \lambda, x_i) = \begin{cases} y_i = 1 \implies \dfrac{e^{\lambda^T x_i}}{1 + e^{\lambda^T x_i}} = \dfrac{1}{1 + e^{-\lambda^T x_i}} = \dfrac{1}{1 + e^{-y_i \lambda^T x_i}} \\ y_i = -1 \implies 1 - \dfrac{e^{\lambda^T x_i}}{1 + e^{\lambda^T x_i}} = \dfrac{1 + \cancel{e^{\lambda^T x_i}} - \cancel{e^{\lambda^T x_i}}}{1 + e^{\lambda^T x_i}} = \dfrac{1}{1 + e^{\lambda^T x_i}} = \dfrac{1}{1 + e^{-y_i \lambda^T x_i}} \end{cases}$$

$$\boxed{P(Y = y_i \mid \lambda, x_i) = \frac{1}{1 + e^{-y_i \lambda^T x_i}}}$$

$$\lambda^* \in \underset{\lambda}{\operatorname{argmax}}\ \log L(\lambda) = \underset{\lambda}{\operatorname{argmax}}\ \log \prod_{i=1}^{n} \frac{1}{1 + e^{-y_i \lambda^T x_i}}$$

$$= \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} \log\left(\frac{1}{1 + e^{-y_i \lambda^T x_i}}\right)$$

$$= \underset{}{\operatorname{argmin}} \sum_{i=1}^{n} \underbrace{\log\left(1 + e^{-y_i \lambda^T x_i}\right)}_{\text{logistic loss}}$$

$$\lambda_1 \overset{1,0}{x_1} + \lambda_2 \overset{1,0}{x_2} + \dots$$