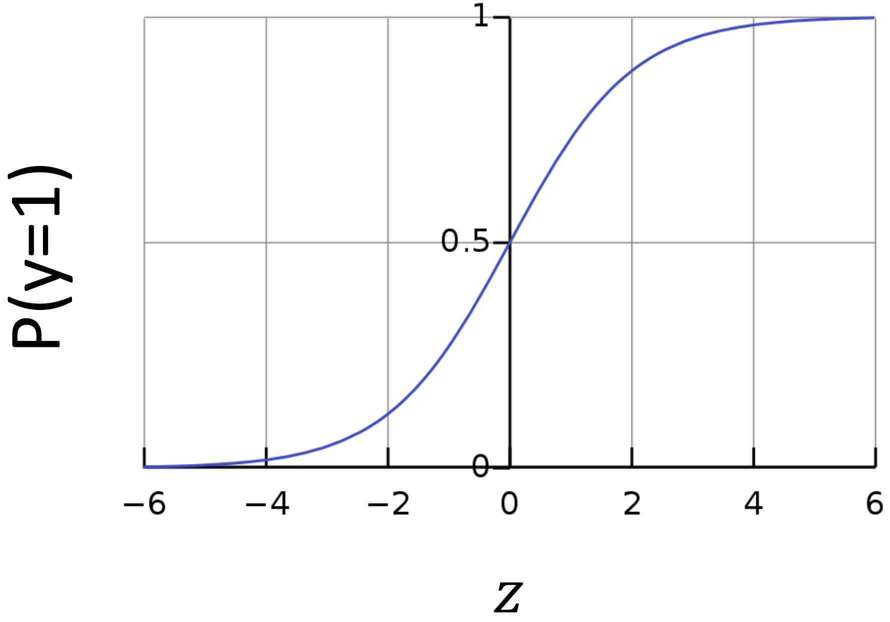
Assessing the coefficients

Logistic Regression Model

$$P(y=1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}}$$

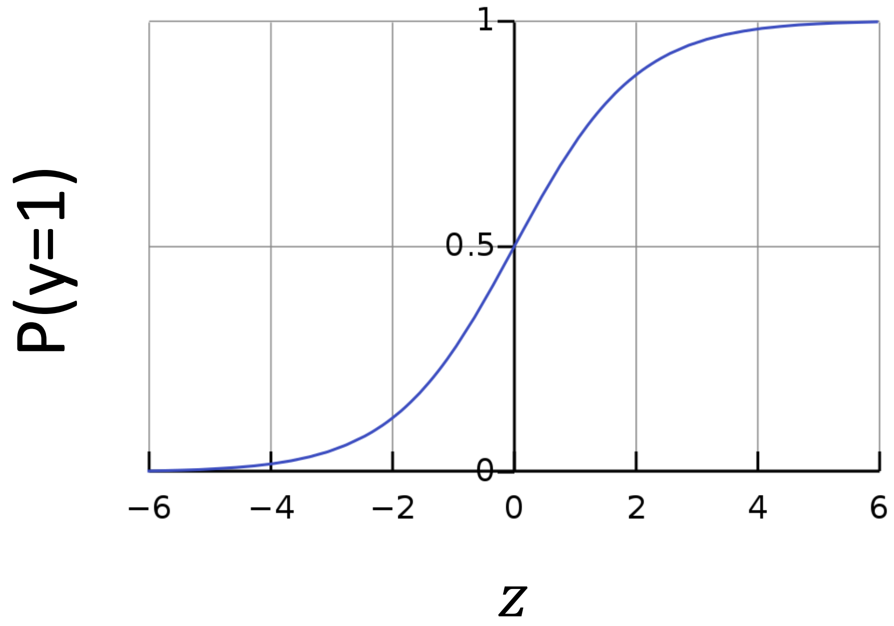
$$P(y=1 \mid X) = \frac{1}{1 + e^{-z}}$$



Logistic Regression Model

$$(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n) > 0$$

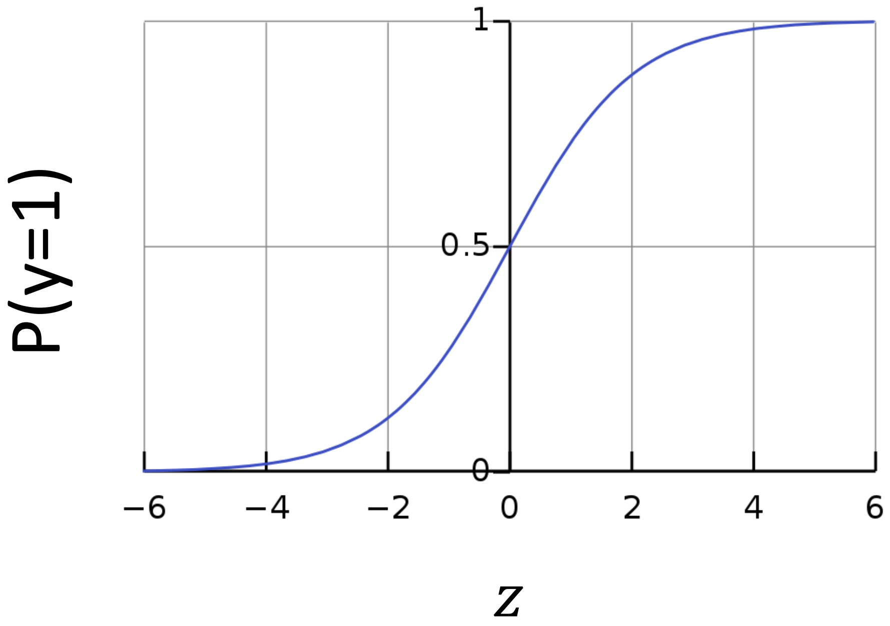
- **β = 0,** the feature does NOT affect the probability of the outcome.
- β > 0 ➔ increase p(y=1)
- β < 0 ➔ decrease p(y=1)



Logistic Regression Model

$$(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n) > 0$$

Is β significantly different from 0?



Wald-test

Null hypothesis: $\beta = 0$

$$z = \frac{\beta}{SE\beta}$$

z ~ normal distribution

Large β or smaller $SE\beta$ indicate that we can trust the coefficient.

Wald-test

Null hypothesis: β = 0

$$z = \frac{\beta}{SE\beta}$$

z ~ normal distribution

When β is large, $SE\beta$ is inflated

➔ z is underestimated

Coefficients error

How can we calculate it?

Coefficients error estimation

- Cross-validation
- Bootstrapping
- Empirically

Statsmodels

$$\hat{\mathrm{se}}_j = \sqrt{((X^T \hat{W} X)^{-1})_{jj}},$$

- X is the predictor matrix
- W is the covariance matrix

https://web.stanford.edu/class/archive/stats/stats200/stats200.1172/Lecture26.pdf

THANK YOU

www.trainindata.com