LOGISTIC REGRESSION (LR)

1. LOGISTIC REGRESSION = LOGISTIC MODEL = LOGIT MODEL
   THIS IS REALLY A CLASSIFICATION SCHEME

2. DIFFERENT WAYS OF EXPRESSING PROBABILITY

O_1, O_2 ARE TWO EVENTS

$p(O_1) = p$

$p(O_2) = q = (1-p)$

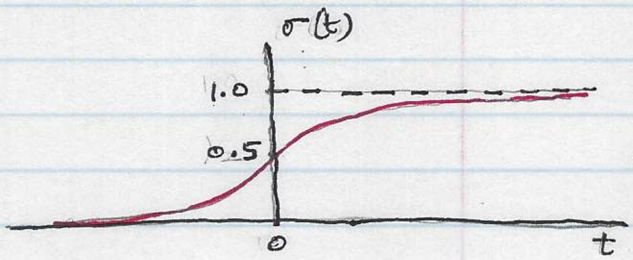
PROBABILITY OF O_1 CAN BE EXPRESSED AS:

- STANDARD PROBABILITY : p ; $p \in [0,1]$
- ODDS : $p/q \triangleq \alpha$; $\alpha \in [0, \infty)$
- LOG ODDS (LOGIT) : $\log\left(\frac{p}{q}\right) \triangleq \gamma$; $\gamma \in (-\infty, +\infty)$

NATURAL LOG

EXAMPLE: IF $p = 0.5$, THEN $\alpha = 1$; $\gamma = 0$

3. LOGISTIC FUNCTION



LOGISTIC FUNCTION $\sigma(t)$, $t \in \mathbb{R}$

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1+e^{-t}}$$

$\sigma(-\infty) = 0$; $\sigma(0) = 0.5$; $\sigma(+\infty) = 1$

$\therefore 0 \le \sigma(t) \le 1 \Rightarrow \sigma(t)$ IS INTERPRETABLE AS PROBABILITY.

4. LOGISTIC REGRESSION ANALYZES RELATIONSHIP BETWEEN A SINGLE (OR MULTIPLE) INDEPENDENT VARIABLE(S) AND A CATEGORICAL DEPENDENT VARIABLE ; AND ESTIMATES THE PROBABILITY OF OCCURRENCE OF ANY EVENT BY FITTING DATA TO A LOGISTIC CURVE.

p = PROBABILITY OF AN EVENT

$\frac{p}{1-p}$ = ODDS OF THE EVENT

$\text{logit}(p) \triangleq \ln\left(\frac{p}{1-p}\right)$ = LOG OF ODD OF EVENT $= \beta_0 + \beta_1 t$; WHERE

t = EXPLANATORY VARIABLE

β_0, β_1 = PARAMETERS OF LR

IF THERE ARE K EXPLANATORY VARIABLES, THEN

$$\text{logit}(p) = \beta_0 + \sum_{i=1}^{k} \beta_i t_i$$

β_i's ARE REGRESSION COEFFICIENTS

& t_i's ARE EXPLANATORY VARIABLES.

5. ADVANTAGES OF LOGISTIC REGRESSION

i) MAKES NO ASSUMPTIONS ABOUT DISTRIBUTIONS OF CLASSES IN FEATURE SPACE

ii) EASILY EXTENDED TO MULTIPLE CLASSES (MULTINOMIAL REGRESSION)

iii) NATURAL PROBABILISTIC VIEW OF CLASS PREDICTIONS

iv) QUICK TO TRAIN

v) VERY FAST AT CLASSIFYING UNKNOWN RECORDS.

vi) GOOD ACCURACY FOR MANY SIMPLE DATA SETS.

vii) RESISTANT TO OVERFITTING.

viii) CAN INTERPRET MODEL COEFFICIENTS AS INDICATORS OF FEATURE IMPORTANCE

<u>DISADVANTAGES OF LOGISTIC REGRESSION</u>

LINEAR DECISION BOUNDARY

6. THE CATEGORICAL PREDICTION CAN BE BASED ON THE COMPUTED ODDS OF A SUCCESS, WITH PREDICTED ODDS ABOVE SOME CHOSEN CUTOFF VALUE BEING TRANSLATED INTO A PREDICTION OF A SUCCESS.

7. ESTIMATION OF REGRESSION COEFFICIENTS

THE REGRESSION COEFFICIENTS ARE USUALLY ESTIMATED USING MAXIMUM LIKELIHOOD ESTIMATION.

IT IS NOT POSSIBLE TO FIND A CLOSED-FORM EXPRESSION FOR THE COEFFICIENT VALUES THAT MAXIMIZE THE LIKELIHOOD FUNCTION, SO THAT AN ITERATIVE PROCESS MUST BE USED INSTEAD; FOR EXAMPLE NEWTON-RAPHSON METHOD.

LET: n = SAMPLE SIZE

$Y = 0, 1$; SPECIFIES THE TWO CATEGORIES

$p = P(Y=1)$

$q = (1-p) = P(Y=0)$

$X_1, X_2, \ldots, X_k$ ARE k PREDICTOR (EXPLANATORY) VARIABLES

$$\text{logit}(p) = \beta_0 + \sum_{j=1}^{k} \beta_j X_j \triangleq A$$

LET $(x_{i1}, x_{i2}, \ldots, x_{ik} ; y_i)$ DENOTE THE VALUES OF $(X_1, X_2, \ldots, X_k ; Y)$ FOR THE i^{th} OBSERVATION, WHERE $1 \le i \le n$.

ASSUME EACH y_i AS THE OUTCOME OF AN INDEPENDENT BERNOULLI RANDOM VARIABLE WITH SUCCESS PROBABILITY p_i ; WE HAVE THE LIKELIHOOD FUNCTION:

$$\prod_{i=1}^{n} p_i^{y_i} (1-p_i)^{1-y_i}$$

$$\ln \frac{p}{1-p} = A \Rightarrow \frac{p}{1-p} = e^{A} \Rightarrow p = \frac{e^{A}}{1+e^{A}} \; ; \; (1-p) = \frac{1}{1+e^{A}}$$

$$p^{y}(1-p)^{1-y} = \frac{e^{Ay}}{(1+e^{A})}$$

$$\prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i} = \frac{\exp\left\{ \sum_{i=1}^{n} y_i \left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \right) \right\}}{\prod_{i=1}^{n} \left\{ 1 + \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \right) \right\}}$$

THE MLE'S OF $(\beta_0, \beta_1, \ldots, \beta_k)$ ARE: $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$.

THESE MAXIMIZE THE LIKELIHOOD FUNCTION.