LOGISTIC REGRESSION

DALPIAZ
STAT 432

So far...

CLASSIFIER

ESTIMATED DISTRIBUTION OF $Y \mid X$

Create $C(x)$ using $\hat{p}_k(x)$

$$\hat{p}_k(x) = \hat{P}[Y = k \mid X = x] \approx$$ Proportion of $y_i = k$ "near" x

↳ KNN (NEIGHBORS)

↳ TREES (NEIGHBORHOODS)

NON-PARAMETRIC MODELS

Now...

A PARAMETRIC METHOD FOR BINARY CLASSIFICATION

BINARY CLASSIFICATION

$$Y = \begin{cases} 1 & \text{"POSITIVE"} \\ 0 & \text{"NEGATIVE"} \end{cases}$$

DEFINE

$$p(x) = P[Y = 1 \mid X = x]$$

OUR FOCUS →

$$1 - p(x) = P[Y = 0 \mid X = x]$$

LOGISTIC REGRESSION

$$\log\underbrace{\left(\frac{p(x)}{1-p(x)}\right)}_{\text{ODDS}} = \underbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p}_{\text{LINEAR COMBINATION OF FEATURES}}$$

$$\underbrace{p(x)}_{P[Y=1 \mid X=x]} = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}{\underbrace{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}_{f(x,\beta)}}$$

LOGISTIC REGRESSION

FUNCTION OF X'S AND β'S

$$Y \mid X \sim \text{BERN}(p(x))$$

COMPARE TO ORDINARY LINEAR REGRESSION

$$Y \mid X \sim N(\underbrace{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}_{\text{FUNCTION OF X'S AND } \beta\text{'S}}, \sigma^2)$$

EXTRA PARAMETER

DEFINE

$$\text{logit}(z) = \log\left(\frac{z}{1-z}\right) \qquad \sigma(z) = \text{logit}^{-1}(z) = \frac{e^{z}}{1+e^{z}} = \frac{1}{1+e^{-z}}$$

$$\text{logit} : [0,1] \to \mathbb{R}$$

$$\sigma : \mathbb{R} \to [0,1]$$

$$\eta(x) = \beta_0 + \beta_1 x_1 + \cdots \beta_p x_p$$

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

$$\text{logit}(p(x)) = \eta(x)$$

$$p(x) = \sigma(\eta(x)) = \frac{e^{\eta(x)}}{1+e^{\eta(x)}} = \frac{e^{\beta_0+\beta_1 x_1 + \cdots + \beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1 + \cdots + \beta_p x_p}}$$

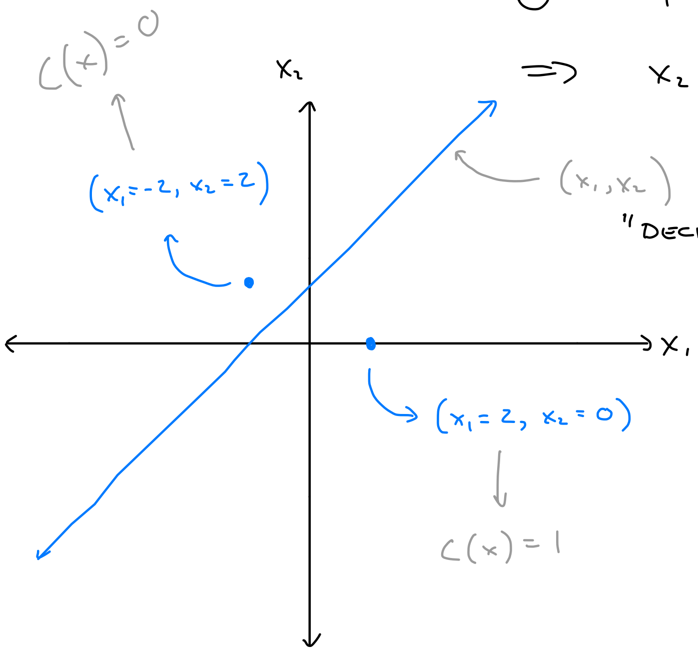
EXAMPLE

$$\log\left(\frac{p(x)}{1-p(x)}\right) = 4 + 2x_1 - 2x_2$$

NOTE $p(x) = 0.5 \iff \eta(x) = 0$

$$0 = 4 + 2x_1 - 2x_2$$

$$\Rightarrow \quad x_2 = 2 + x_1$$



$(x_1, x_2) : p(x) = 0.5$

"DECISION BOUNDARY"

$$P(x_1=2, x_2=0) = \frac{1}{1+e^{-(4+4+0)}} = 0.9996$$

$$P(x_1=-2, x_2=2) = \frac{1}{1+e^{-(4-4-4)}} = 0.01799$$

Fitting Logistic to Data

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x$$

x_i	y_i	$p(x_i)$
2	1	
3	1	
1	1	
3	1	
5	1	
4	0	
5	0	
6	0	
7	0	
6	0	

Should be "large"

Should be "small"

SEQUENCE : 1, 1, 0

PROBABILITY : $p(x_1) \cdot p(x_2) \cdot (1 - p(x_3))$

Conditional Likelihood

$$\mathcal{L}(\beta_0, \beta_1) = \prod_{i=1}^{n} P[Y_i = y_i \mid X_i = x_i]$$

MAXIMIZE

$$\mathcal{L}(\beta_0, \beta_1) = \prod_{i=1}^{n} P[Y_i = y_i \mid X_i = x_i] = \prod_{i=1}^{n} p(x_i)^{y_i} (1 - p(x_i))^{1-y_i}$$

$$\log \mathcal{L}(\beta_0, \beta_1) = \underbrace{\sum_{i=1}^{n} y_i \log(p(x_i))}_{\text{CLASS 1}} + \underbrace{\sum_{i=1}^{n} (1-y_i) \log(1-p(x_i))}_{\text{CLASS 0}}$$

$$= \sum_{i=1}^{n} \log(1-p(x_i)) + \sum_{i=1}^{n} y_i \log\left(\frac{p(x_i)}{1-p(x_i)}\right)$$

$$= \sum_{i=1}^{n} \log\left(1 - \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right) + \sum_{i=1}^{n} y_i (\beta_0 + \beta_1 x_i)$$

$$= -\sum_{i=1}^{n} \log\left(1 + e^{\beta_0+\beta_1 x_i}\right) + \sum_{i=1}^{n} y_i (\beta_0 + \beta_1 x_i)$$

$$\log \mathcal{L}(\beta_0, \beta_1) = -\sum_{i=1}^{n} \log\left(1 + e^{\beta_0 + \beta_1 x_i}\right) + \sum_{i=1}^{n} y_i (\beta_0 + \beta_1 x_i)$$

$$\boxed{\begin{aligned} \frac{d}{d\beta_0} \log \mathcal{L}(\beta_0, \beta_1) &= -\sum_{i=1}^{n} \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} + \sum_{i=1}^{n} y_i = 0 \\ \frac{d}{d\beta_1} \log \mathcal{L}(\beta_0, \beta_1) &= -\sum_{i=1}^{n} x_i \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} + \sum_{i=1}^{n} x_i y_i = 0 \end{aligned}}$$

- NO CLOSED FORM SOLUTION
- USE NUMERIC OPTIMIZATION
  - NEWTON'S METHOD
  - IRLS
  - GRADIENT DESCENT

OR...

LOGISTIC REGRESSION IN R

FITTING : glm(formula, data, family = "binomial")

↳ MAKE SURE RESPONSE IS A FACTOR

↳ FIRST LEVEL → $Y=0$

↳ SECOND LEVEL → $Y=1$

"PREDICTING" : predict()

↳ type = "link" → $\hat{\eta}(x)$

↳ type = "response" → $\hat{p}(x)$

coef() → $\hat{\beta}_0, \hat{\beta}_1, \ldots$