

LOGISTIC REGRESSION

DALPIAZ
STAT 432

So FAR ...

CREATE $C(x)$ USING $\hat{P}_k(x)$

CLASSIFIER

ESTIMATED DISTRIBUTION OF $Y|X$

$$\hat{P}_k(x) = \hat{P}[Y=k | X=x] \approx \text{PROPORTION OF } y_i = k \text{ "NEAR" } x$$

↳ 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

OUR FOCUS \rightarrow $\rho(x) = P[Y=1 | X=x]$
 $1 - \rho(x) = P[Y=0 | X=x]$

LOGISTIC REGRESSION

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

↑
ODDS

LINEAR COMBINATION OF FEATURES

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

↑
 $P[Y=1 | X=x]$

$f(x, \beta)$

LOGISTIC REGRESSION

FUNCTION OF X'S AND β 'S

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

COMPARE TO ORDINARY LINEAR REGRESSION

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

FUNCTION OF
X'S AND β 'S

EXTRA PARAMETER

DEFINE

$$\text{logit}(\zeta) = \log\left(\frac{\zeta}{1-\zeta}\right)$$

$$\sigma(\zeta) = \text{logit}^{-1}(\zeta) = \frac{e^{\zeta}}{1+e^{\zeta}} = \frac{1}{1+e^{-\zeta}}$$

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

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

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

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

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

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

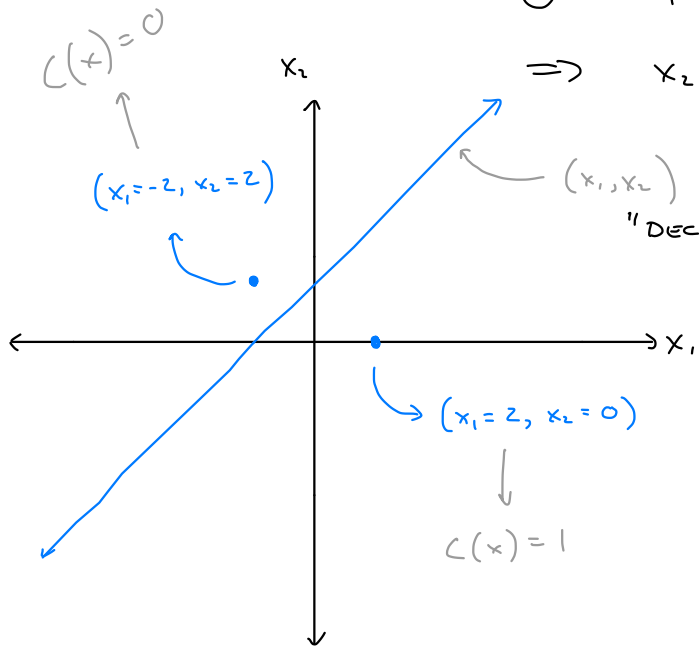
EXAMPLE

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

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

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

$$\Rightarrow x_2 = 2 + x_1$$



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

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

FITTING LOGISTIC TO DATA

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

<u>X_i</u>	<u>y_i</u>	<u>$p(x_i)$</u>
2	1	}
3	1	
1	1	
3	1	
5	1	
4	0	}
5	0	
6	0	
7	0	
6	0	

SHOULD BE
"LARGER"

SHOULD BE
"SMALL"

SEQUENCE : 1, 1, 0

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

CONDITIONAL LIKELIHOOD

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

MAXIMIZE ↗

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

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

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

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

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

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

$$\frac{\partial}{\partial \beta_0} \log \mathcal{L}(\beta_0, \beta_1) = -\sum_{i=1}^n \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} + \sum_{i=1}^n y_i = 0$$

$$\frac{\partial}{\partial \beta_1} \log \mathcal{L}(\beta_0, \beta_1) = -\sum_{i=1}^n x_i \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} + \sum_{i=1}^n x_i y_i = 0$$

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

OR...

LOGISTIC REGRESSION IN R

FITTING: $\text{glm}(\text{formula}, \text{data}, \text{family} = \text{"binomial"})$

↳ MAKE SURE RESPONSE IS A FACTOR

↳ FIRST LEVEL $\rightarrow Y=0$

↳ SECOND LEVEL $\rightarrow Y=1$

"PREDICTING": $\text{predict}()$

↳ $\text{type} = \text{"link"} \rightarrow \hat{\eta}(x)$

↳ $\text{type} = \text{"response"} \rightarrow \hat{p}(x)$

$\text{coef}() \rightarrow \hat{\beta}_0, \hat{\beta}_1, \dots$