CS 307

SPRING 2024

DALPIAZ

LINEAR CLASSIFICATION

LOGISTIC REGRESSION

????

So far...

Create $C(x)$ using $\hat{p}_k(x)$ estimating $Y|X$

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

↳ KNN neighbors

↳ Tree neighborhoods

Nonparametric

Now...

A parametric method for

<u>Binary</u> classification

Binary Classification

$$Y = \begin{cases} 1 & \text{"positive"} \\ 0 & \text{"negative"} \end{cases}$$

Define

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

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

LOGISTIC REGRESSION

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

ODDS

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

$$f(x, \beta)$$

KNOW LEARN

Logistic Regression

$$Y \mid X \sim \text{Bern}(\overbrace{p(x)}^{f(x,\beta)})$$

Compare to Ordinary Linear Regression

$$Y \mid X \sim N(\underbrace{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}_{\text{linear combo}}, \sigma^2)$$

Parameter

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

DEFINE

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

SOME INPUT

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

SIGMOID FUNCTION

INVERSE LOGIT

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

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

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

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

$$\boxed{\operatorname{logit}(p(x)) = \eta(x)}$$

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

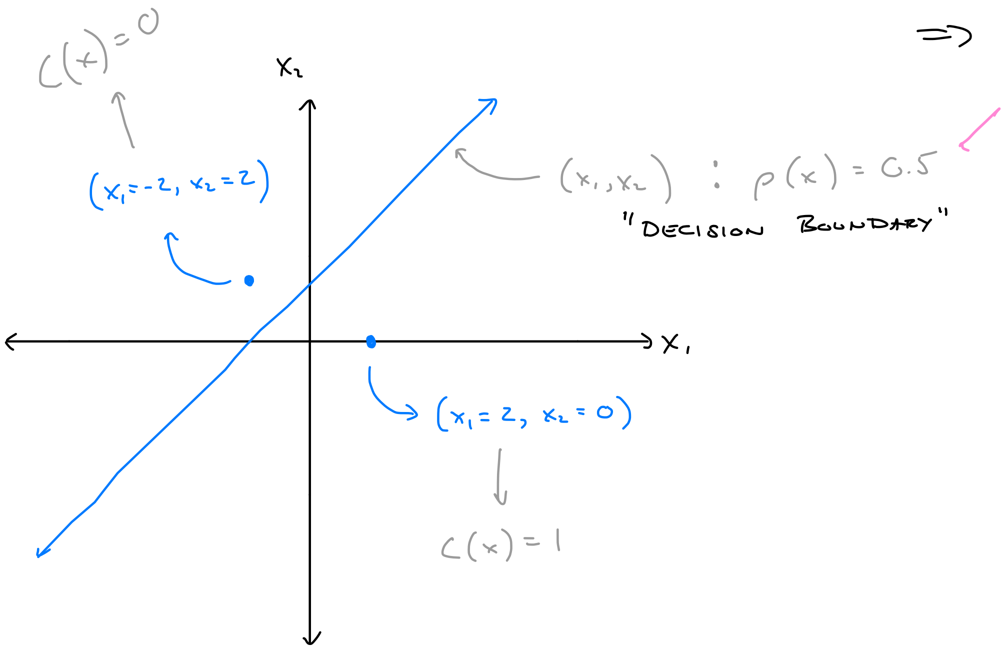
EXAMPLE

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \underbrace{4}_{\beta_0} + \underbrace{2}_{\beta_1}x_1 - \underbrace{2}_{\beta_2}x_2$$

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

$$0 = 4 + 2x_1 - 2x_2 = \eta(x)$$

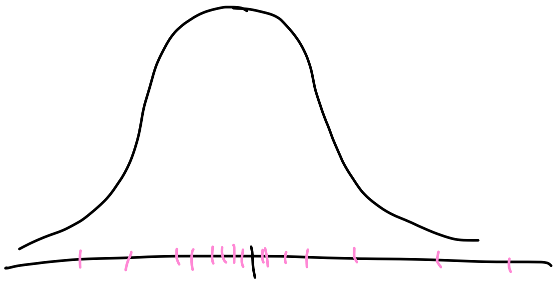
$$\Rightarrow \quad x_2 = 2 + x_1$$

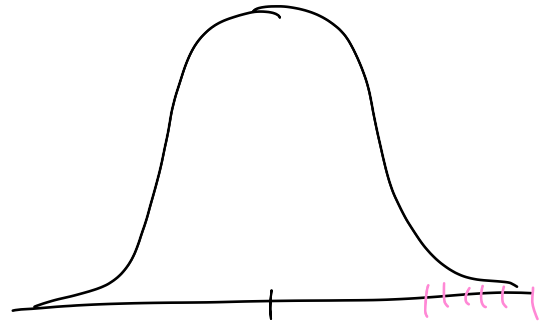


$$P(x_1=2, x_2=0) = \frac{1}{1+e^{-(4+4\cdot 0)}} = 0.9996$$

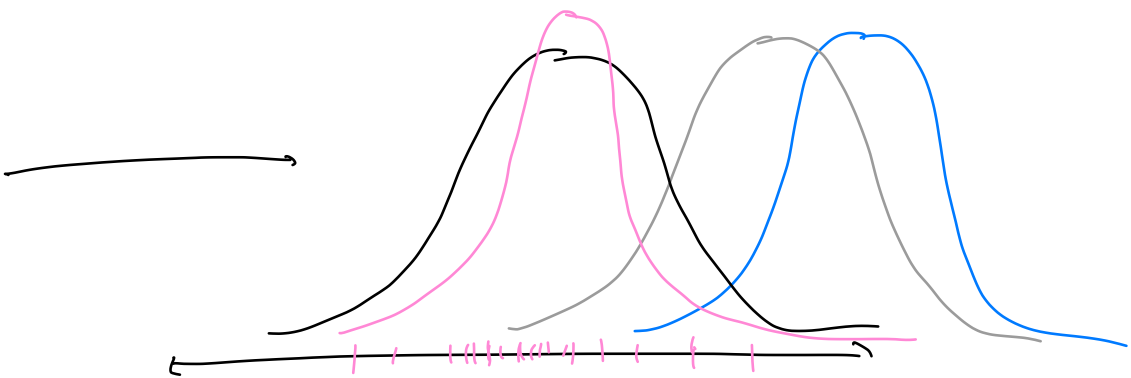
$$P(x_1=-2, x_2=2) = \frac{1}{1+e^{-(4-4-4)}} = 0.01799$$

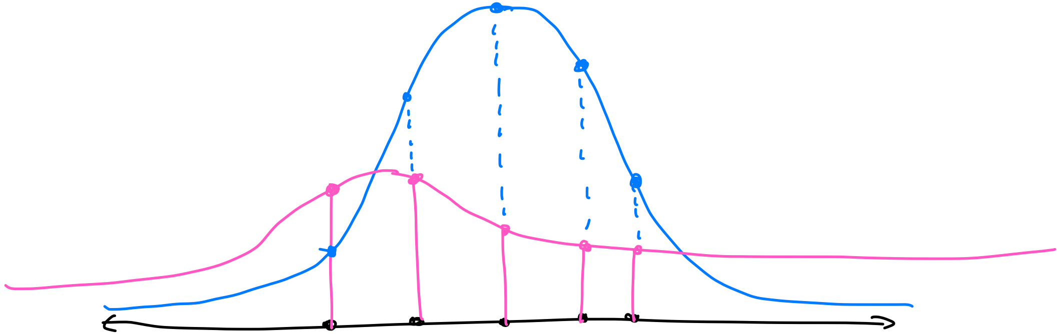
A NOTE ABOUT LIKELIHOOD





μ, σ^2





PDF of $N(\mu, \sigma^2)$

$$\mathcal{L}(\mu, \sigma \mid x) = \prod_{i=1}^{n} f(x_i \mid \mu, \sigma)$$

Fitting Logistic to Data

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"

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

y_1 y_2 y_3

Sequence : 1, 1, 0

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

$$\boxed{p(x)^{y}\,(1-p(x))^{1-y}}$$

Conditional Likelihood

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

Maximize

$f(x, \beta)$

Known

Learn

LIKELIHOOD

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

$y_i = 1$

$1 - y_i = 0$

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

LOG-LIKELIHOOD

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

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

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

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

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

- NO CLOSED FORM SOLUTION
- USE NUMERIC OPTIMIZATION
  - NEWTON'S METHOD
  - ETC

OR...

Logistic Regression in Python

sklearn.linear_model.LogisticRegression

- .fit
- .predict
- .predict_proba

Not logistic regression by default...

Must set

penalty = None