LINEAR CLASSIFICATION

CS 307

LOGISTIC REGRESSION

????

So far...

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

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

↳ KNN neighbors

↳ Tree neighborhoods

Nonparametric

Now...

A parametric method for

Binary classification

Binary Classification

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

Define

Notation

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

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

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

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

$$f(x, \beta)$$

KNOWN

LEARNED

LOGISTIC REGRESSION

$f(x, \beta)$

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

COMPARE TO ORDINARY LINEAR REGRESSION

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

ADDITIONAL PARAMETER

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

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

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

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

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

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

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

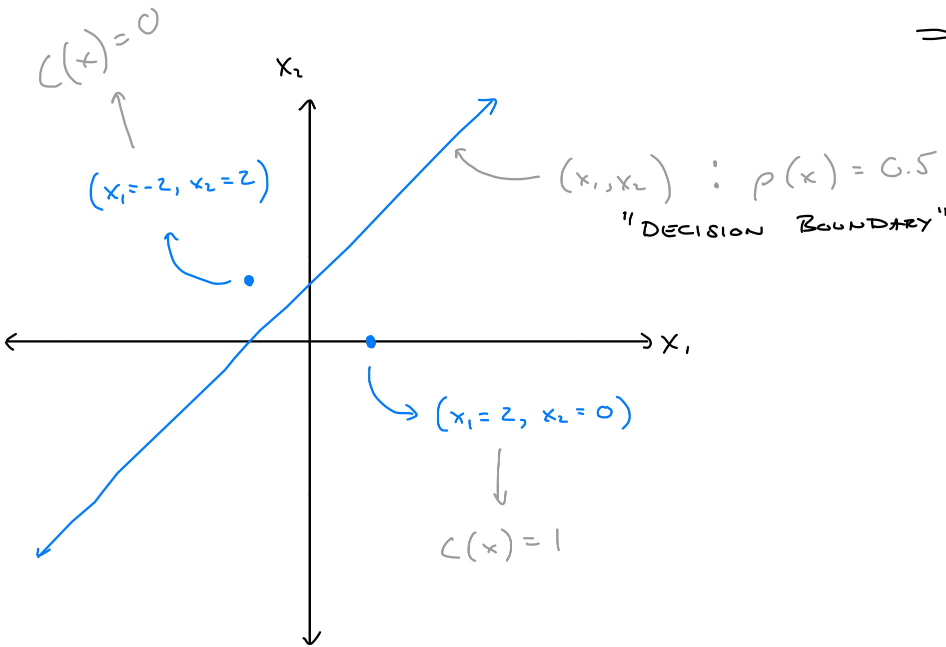
Example

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

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

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

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



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

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

Logistic Regression in Python

sklearn.linear_model.LogisticRegression

- .fit
- .predict
- .predict_proba

Not logistic regression by default...

Must set

penalty = None