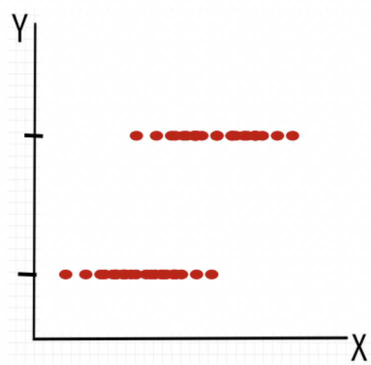
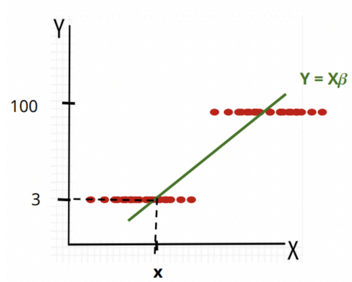
CAS CS 506
Lec 19

Logistic Regression

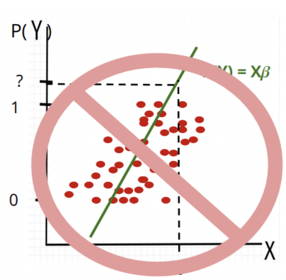
1. Logistic Regression
    a. What if y_i is categorical? Can we use a linear function to predict y_i?
    b. Assume we have 2 classes



    c. The linear model will look like

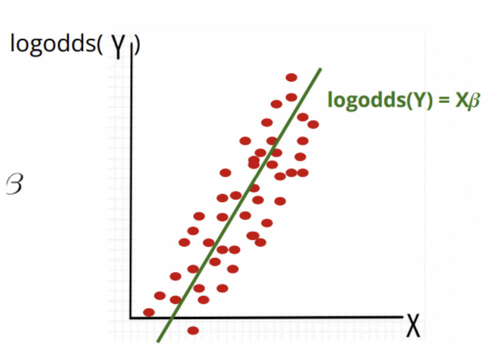


    d. This is not good
    e. The numerical values associated with the class are arbitrary numbers. A model based on these numbers would be meaningless
    f. So we should NOT model the class itself with a linear model
    g. Notice that a linear function will predict a continuum of values. So we should find an interpretation / transformation of the class that is continuous for us to predict
    h. Can we use the probability of belonging to a given class as a proxy for how confidently we can classify a given point?



i. So it's not just a continuum of values - the range of values needs to be (-inf, +inf)
j. Define the odds = p / 1 - p where p = P(Y = class 1 | X)
k. Now the range of XB_{LS} is [0, +inf)
l. In order to get (-inf, inf), let's take the log of the odds! This is also convenient numerically because in the odds format, tiny variations in p have large effects on the odds
m. Our goals is to fit a linear model to the log-odds of being in one of our classes (in the 2-class case) i.e.

$$\log\left(\frac{P(Y=1|X)}{1-P(Y=1|X)}\right) = X\beta$$

n. 

o. Predict this model by

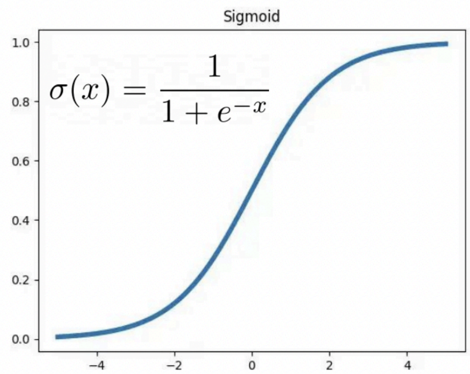
DECISION RULE:
IF P(Y=1|X) > ½ THEN 1 ELSE 0

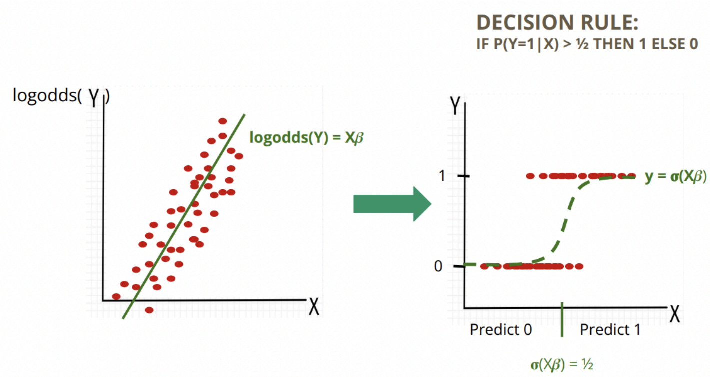
p. Suppose we have such a model. How do we recover the P(Y = 1 | X)?

$$\begin{aligned}
\log\left(\frac{P(Y=1|X)}{1-P(Y=1|X)}\right) &= \alpha + \beta X \\
\frac{P(Y=1|X)}{1-P(Y=1|X)} &= e^{\alpha+\beta X} \\
\frac{P(Y=1|X)}{1-P(Y=1|X)} + 1 &= e^{\alpha+\beta X} + 1 \\
\frac{P(Y=1|X)}{1-P(Y=1|X)} &= e^{\alpha+\beta X} + 1 \\
P(Y=1|X) &= \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}
\end{aligned}$$

i. The function we apply to our probability to obtain the log odds is called the logit function. The function used to retrieve our probability from the log odds is called logit^-1 or sigmoid

q. $$\log\left(\frac{P(Y=1|X)}{1-P(Y=1|X)}\right) = \alpha + \beta X \qquad P(Y=1|X) = \sigma(\alpha + \beta x)$$



r. 

2. What does the decision boundary look like?
    a. 
    b. Decision boundary is where P(Y = 1 | X) = ½
    c. $$P(Y=1|X) = \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}$$
    d. Decision boundary is where e^(wx + b) = 1
    e. Decision boundary is where wx + b = 0
3. Maximum Likelihood Estimator
    a. How do we learn our model? I.E. the a and B parameters

b. We know:

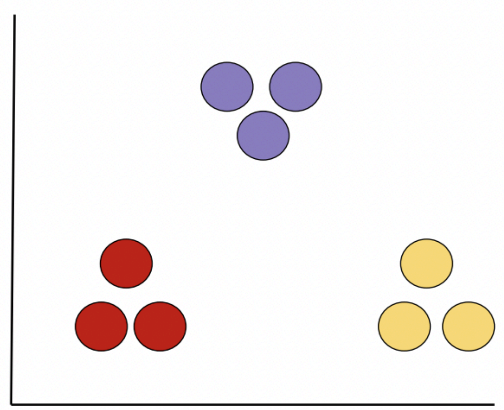
$$P(y_i|x_i) = \begin{cases} \sigma(\alpha + \beta x_i) & \text{if } y_i = 1 \\ 1 - \sigma(\alpha + \beta x_i) & \text{if } y_i = 0 \end{cases}$$

$$P(y_i|x_i) = \sigma(\alpha + \beta x_i)^{y_i}(1 - \sigma(\alpha + \beta x_i))^{1-y_i}$$

c. So we define the probability of having seen the data we saw:

$$\begin{aligned} L(\alpha, \beta) &= \prod_{i=1}^{n} P(y_i|x_i) \\ &= \prod_{i=1}^{n} \sigma(\alpha + \beta x_i)^{y_i}(1 - \sigma(\alpha + \beta x_i))^{1-y_i} \end{aligned}$$

4. What if there are 3 classes

a. 

$$\log\left(\frac{P(Y=0|X)}{P(Y=2|X)}\right) = \beta_0 X$$

$$\log\left(\frac{P(Y=1|X)}{P(Y=2|X)}\right) = \beta_1 X$$

b. $P(Y=2|X) = 1 - (P(Y=1|X) + P(Y=0|X))$

$$P(Y=0|X) = \frac{e^{\beta_0 X}}{1 + e^{\beta_0 X} + e^{\beta_1 X}}$$

$$P(Y=1|X) = \frac{e^{\beta_1 X}}{1 + e^{\beta_0 X} + e^{\beta_1 X}}$$

c. $$P(Y=2|X) = \frac{1}{1 + e^{\beta_0 X} + e^{\beta_1 X}}$$

d.