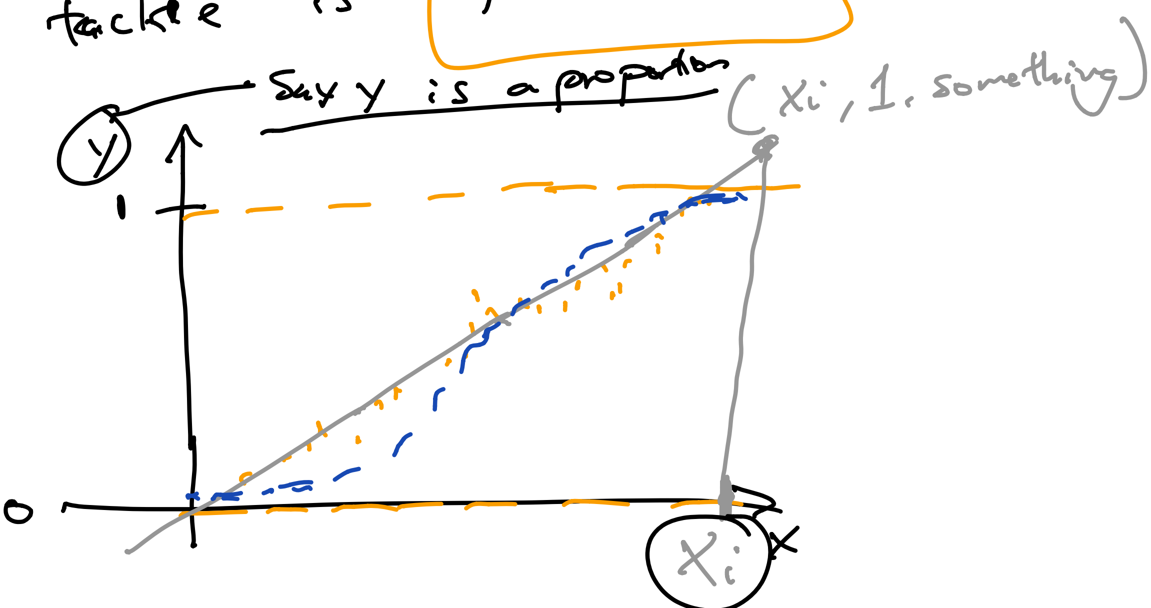
Generalized Linear Models (GLMs) (Chap 8).

- Generalization of OLS specialized for some kinds of y variables.

So far $y \in \mathbb{R}$ or $y \in (-\infty, \infty)$. If we consider instead that y is in some subset of $\mathbb{R}$, we can improve our OLS results by tailoring OLS to the problem on-hand.

- The first of these subsets that we'll learn to tackle is $\boxed{y \in [0, 1]}$

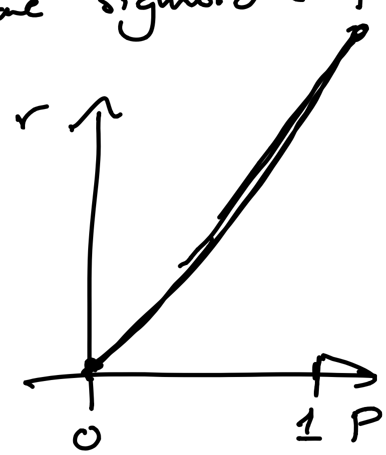
Say y is a proportion



- Let's do something to turn the $(-\infty, \infty)$ domain into $[0, 1]$, (and also have sigmoid shape)

Let:

$$r = \frac{p}{1-p} \quad \text{(odds ratio)}$$



This does the trick. However, it has the disadvantage of growing too quickly. To control the rate at which it grows/shrinks, let's take the log.

$$\lambda = \log\left(\frac{p}{1-p}\right) \quad \text{(logit function)}$$

- If we apply this transformation to each observation p_i, we get

$$\lambda_i = \log\left(\frac{p_i}{1-p_i}\right) \in (-\infty, \infty) \quad \forall i$$

- Now, we can regress λ_i using OLS (yay!)

For the 1-dim case:

$$\lambda_i = \log\left(\frac{p_i}{1-p_i}\right) = \underbrace{\beta_0 + \beta_1 \cdot x_i + \varepsilon_i}_{OLS}$$

- Use MLE to get $\hat{\beta}_0, \hat{\beta}_1 \Rightarrow \hat{\lambda}_i$ and get

$$\boxed{\hat{\lambda}(x) = \hat{\beta}_0 + \hat{\beta}_1 \cdot x}$$

- Finally, we can get $\hat{p}_i$ by transforming back...

$$\hat{\lambda}_i = \log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) \to e^{\hat{\lambda}_i} = \frac{\hat{p}_i}{1-\hat{p}_i}$$

$$\to e^{\hat{\lambda}_i} - \hat{p}_i \cdot e^{\hat{\lambda}_i} = \hat{p}_i \Rightarrow e^{\hat{\lambda}_i} = \hat{p}_i + \hat{p}_i e^{\hat{\lambda}_i}$$

$$= p_i(1 + e$$

$$\Rightarrow \hat{p}_i = \frac{e^{\hat{\lambda}_i}}{(1+e^{\hat{\lambda}_i})}$$

$$= \frac{e^{\hat{\lambda}_i}/e^{\hat{\lambda}_i}}{\left(\frac{1+e^{\hat{\lambda}_i}}{e^{\hat{\lambda}_i}}\right)}$$

$$= \frac{1}{(1+e^{-\hat{\lambda}_i})}$$

$$\Rightarrow \hat{p}_i = \frac{1}{(1+e^{-(\hat{\beta}_0+\hat{\beta}_1 \cdot x_i)})}$$

$$\hat{p}(x) = \left[1 + e^{-(\hat{\beta}_0+\hat{\beta}_1 x_i)}\right]^{-1}$$

Inverse logit function.

We can now write a regression about proportions in same way we write our usual linear regression.

$\Rightarrow$ Now, I'd like to present you with an issue:

Y	X
0	3.2
1	4.1
0	7.1

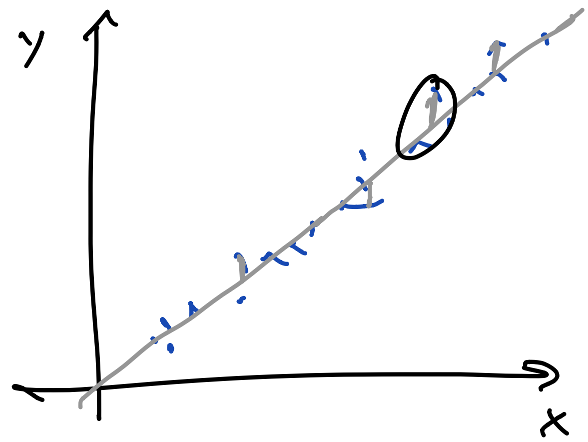
② Our usual objective (loss) function

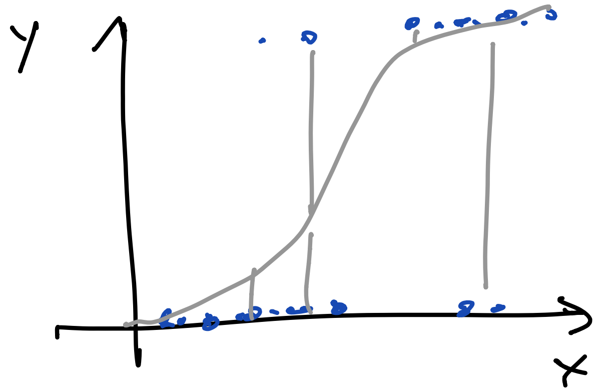
$$\min_{\beta} \|\underline{Y} - \underline{X} \cdot \beta\|^2$$

1	3.1
0	8.9
0	9.0
0	7.7
1	5.0

① $\boxed{\text{Logit breaks !!}}$

... isn't particularly suited for this problem.





- Specifically, thinking about hit or miss situations, It'd be better to have a loss function that did not increase when we get the right prediction and increased if we got the wrong prediction.

$$\underbrace{D}_{\text{Deviance}} = \underbrace{p_i}_{\text{observed}} \cdot \underbrace{\log\left(\frac{p_i}{\hat{p}_i}\right)}_{\text{If obs = 1 and prediction}} + (1-p_i)\cdot \underbrace{\log\left(\frac{1-p_i}{1-\hat{p}_i}\right)}_{\text{If } p_i = 1 \text{ and } \hat{p}_i \to 0^+ \text{ then, large}}$$

= 1

then = 0

penalty.

As usual, to apply this to repeated iid sampling we can get a deviance function for the whole dataset.

$$D(\hat{p}_i | p_i) = n_i \cdot \left[p_i \cdot \log\left(\frac{p_i}{\hat{p}_i}\right) + (1-p_i)\log\left(\frac{1-p_i}{1-\hat{p}_i}\right)\right]$$

Example. Dose - response

- If large dose is administered, drug  increases mortality on test mice

- 11 groups of 10 mice each. Different doses administered to each group.

y_i = # of mice dying in group i

p_i = prop. of deaths in group i

$p_i = y_i / 10$

$$y_i \overset{ind}{\sim} \text{Bin}(10, \underbrace{p_i}_{\text{True death rate of group } i.})$$

- Of course, we can get MLE estimates

for each group $\hat{p}_i = \frac{y_i}{10}$, but using all groups gives us more information, as we can leverage results of other groups.