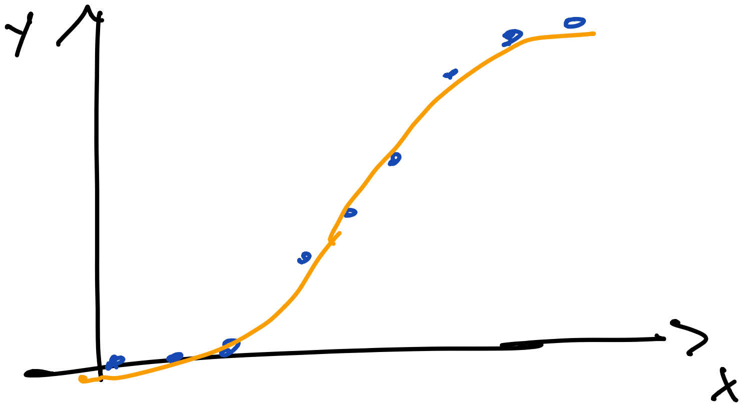
- Second Logistic Regression Example.



Previously ...

-> our $\hat{y}$ was a probability (mortality rate).

Now, we will use logistic reg to <u>classify</u> data.

→ 1 (Spam)

→ 0 (Good email)

① we'll get $\hat{y}_i$'s $\in (0, 1)$ using Logistic Reg.

② we'll compare $\hat{y}_i$'s vs. a <u>threshold</u> (we'll use 0.5).

Rule : $\hat{y}_i \geq$ threshold → 1 Spam

$\hat{y}_i <$ threshold → 0 Good

. we'll work with a medium-dimensional X. ~57 vars.

- In <u>linear models</u>, we need to turn our <u>categorical</u> variables into a set of dummy variables.

β_{vac} work β_{Fam}

Spam	Intercept	Topic		Topic'		Vac_dum	Work_dum	Fam_dum
1	1	Vacation	→	1	→	1	0	0
0	1	Work		2		0	1	0
1	1	Family		3		0	0	1
1	1	⋮		⋮		⋮	0	⋮
0	1							
1	1							
0	⋮							
⋮								

All entries = 1

These are linearly dependent with

$$\begin{array}{c|cccc} & c_1 & c_2 & c_3 & c_4 \\ \hline 1 & & & & \\ 2 & & & & \\ 3 & & & & \\ 4 & & & & \end{array}$$

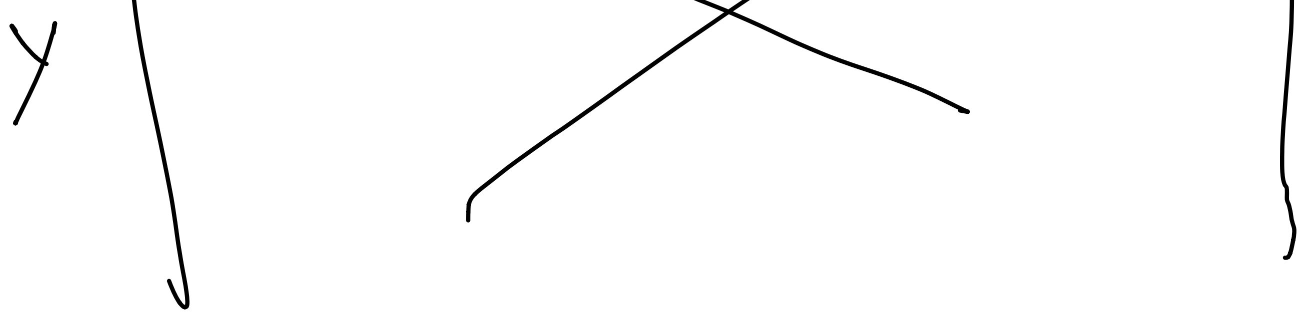
and c_2 and c_3 are linearly dependent.

Max rank this matrix can have is 3.

=> If matrix is not full rank, it is not invertible.

→ $(X^T X)^{-1}$ doesn't exist!!

Y | X_{cont1} X_{cont2} ... X_{contp1} $xd_1, xd_2, \dots xd_{p2}$ |



Generalized Linear Models

- Logistic regression is a special case of a more general construction,
  - → we used the logit function as a "link" between our data and the capabilities of OLS.
  - → Further, we defined the Deviance function, which was a loss function more directly tailored for Logistic Reg.

Distrib. Family	Link Function	Deviance
Binomial	$\boxed{\lambda = \log\left(\frac{p}{1-p}\right)}$	sampling for the whole "dataset," $D(\hat{p}_i, p_i) = n_i \cdot \left[p_i \cdot \log\left(\frac{p_i}{\hat{p}_i}\right) + (1-p_i)\log\left(\frac{1-p_i}{1-\hat{p}_i}\right)\right]$
Poisson	$\lambda = \log(\mu)$	$2\mu\left[\left(\frac{\hat{\mu}}{\mu} - 1\right) - \log\frac{\hat{\mu}}{\mu}\right]$
Normal with σ^2 known	$\lambda = \frac{\mu}{\sigma}$	$\left(\frac{\hat{\mu} - \mu}{\sigma}\right)^2$
Gamma	$\lambda = \frac{-1}{\sigma}$	$2\nu\left[\left(\frac{\hat{\sigma}}{\sigma} - 1\right) - \log\frac{\hat{\sigma}}{\sigma}\right]$

V known

General Definition of Deviance

$$D(f_1, f_2)$$

$$= \int_{y} f_1(y) \log\left\{\frac{f_1(y)}{f_2(y)}\right\} dy$$

If we find the values for $\hat{p}_i$, $\hat{\mu}$ or $\hat{\delta}$ that minimize this func. $\Rightarrow$ $\hat{\mu}^{MLE}$, $\hat{p}_i^{MLE}$, $\hat{\delta}^{MLE}$.

Hoeffding's Lemma

For Binomial (Logistic Reg).

$$\lambda_i = \log\left(\frac{p_i}{1-p_i}\right) = X^T \beta \rightarrow \hat{\beta}$$

$$\begin{bmatrix} | & x_1 & x_2 \\ | & | & | \\ | & | & | \end{bmatrix}$$

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

Poisson Regression

- After OLS and Logistic Reg, Poisson Reg is the <u>third</u> most used linear model.

$$y_i \overset{ind}{\sim} Poi(\mu_i)$$

Using our link function, let $\lambda_i = \log \mu_i$

we will assume $\underline{\lambda} = X\beta$

$$\underline{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}$$

$\rightarrow$ Let's jump to an example

$r \rightarrow$

$m \uparrow$	1	2	3	4	5	...	15
18	1	6	6	1	- - —		
$\vdots$	3	2					
$\vdots$							
3							
2							
1							

# of counts.