



GENES
20000
CELLS 1000

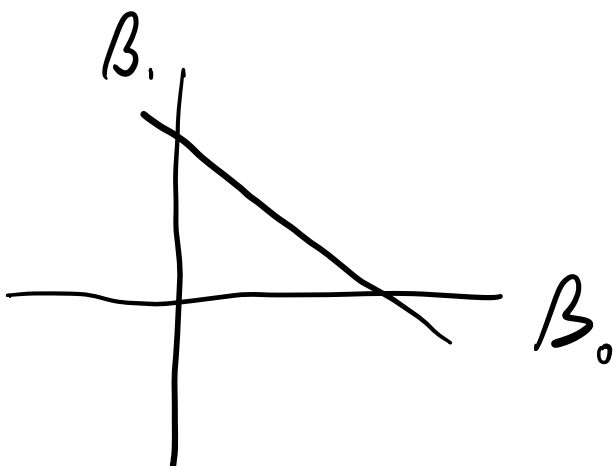
BACTERIAL
SPECIES $\sim 10^4$

PATENTS 10

RIDGE REGRESSION

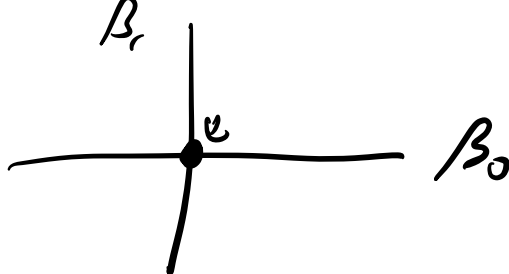
^ SAME

MINIMIZE $SSR + \lambda \sum_i \beta_i^2$



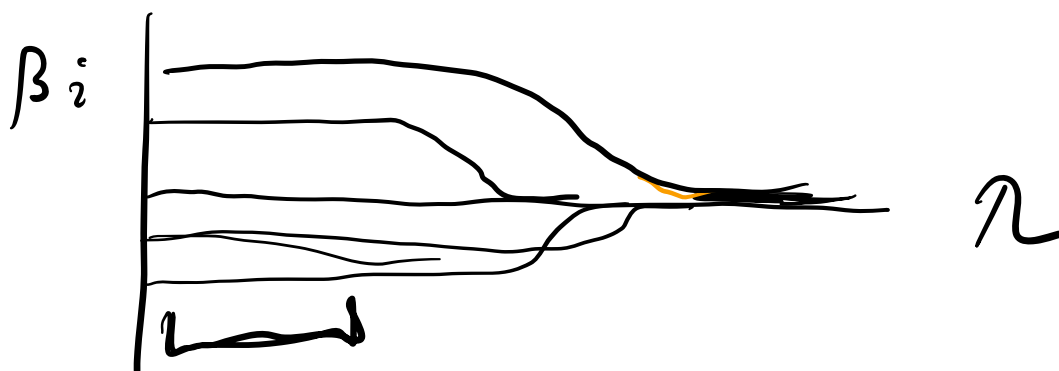
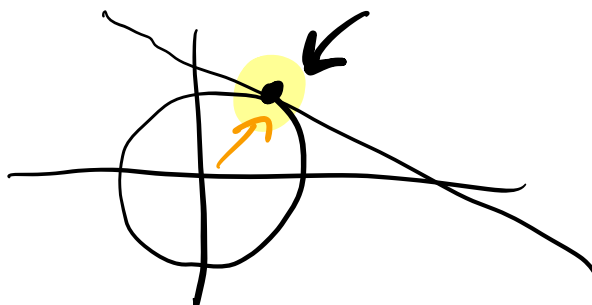
$\beta_0^2 + \beta_1^2 = \text{CONST}$

$$\lambda \rightarrow \infty$$



$$\lambda \rightarrow 0$$

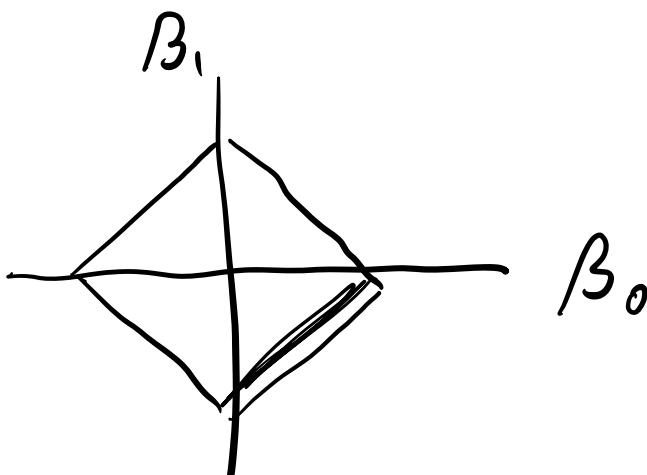
BUT NOT
 $\lambda = 0$

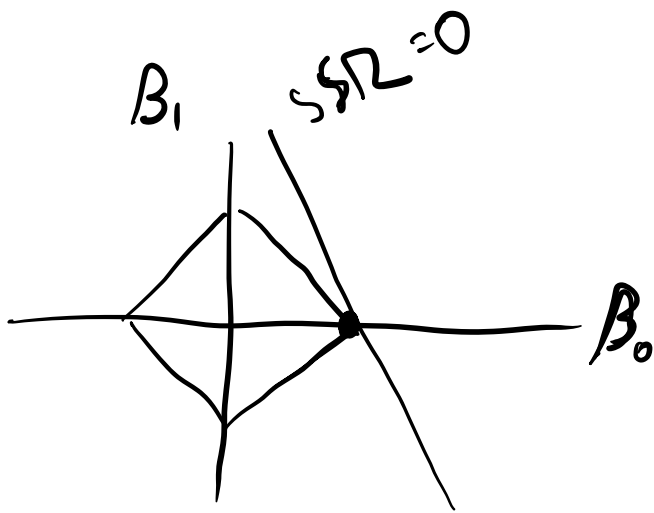


LASSO REGRESSION

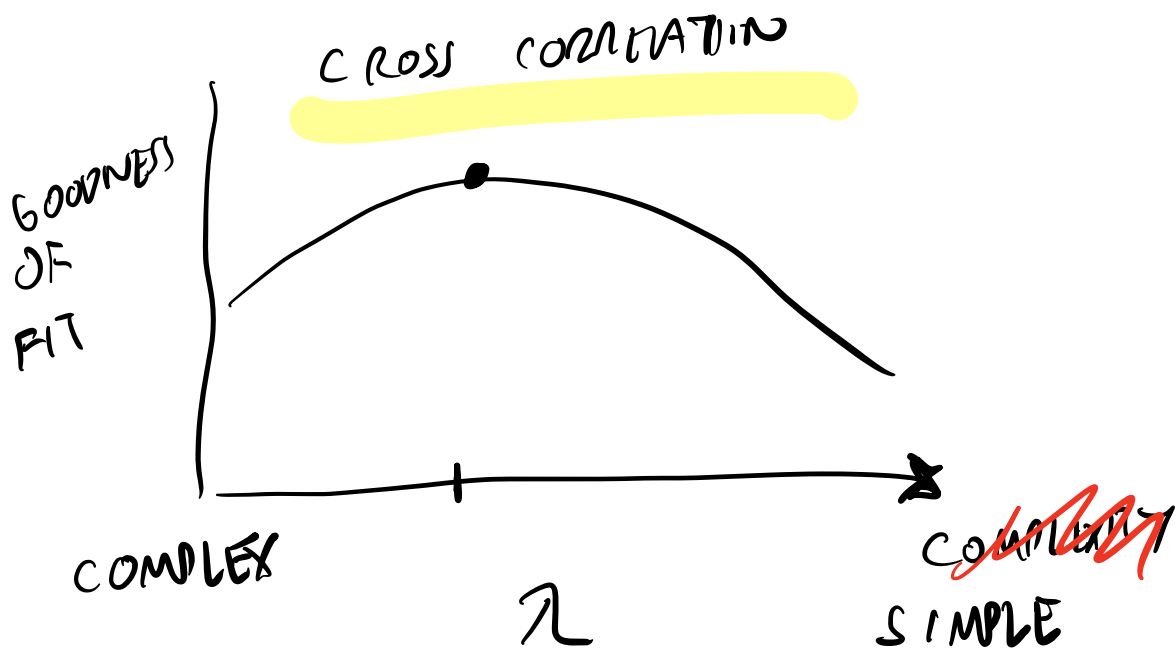
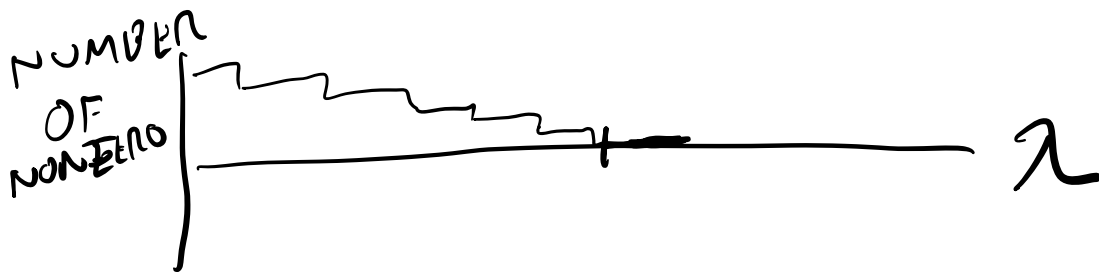
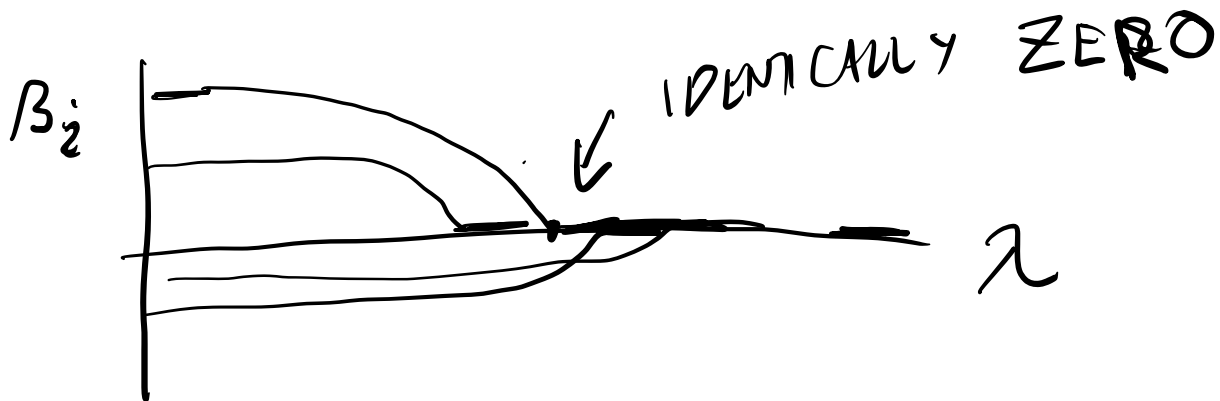
$$\text{MINIMIZE } SSR + \lambda \sum_i | \beta_i |$$

$$| \beta_0 | + | \beta_1 | = C$$





$\lambda \rightarrow 0$



GENERALIZED LINEAR MODEL

PATIENTS WITH / WITHOUT
DIABETES (YES/NO)

"Y"

X_1 GLUCOSE LEVEL IN BLOOD

X_2 INSULIN LEVEL IN BLOOD

\vdots

p_i - PROB THAT PATIENT i HAS DIABETES

$$p_i = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots}}$$

$$\log \left(\frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots$$

LOG ODDS