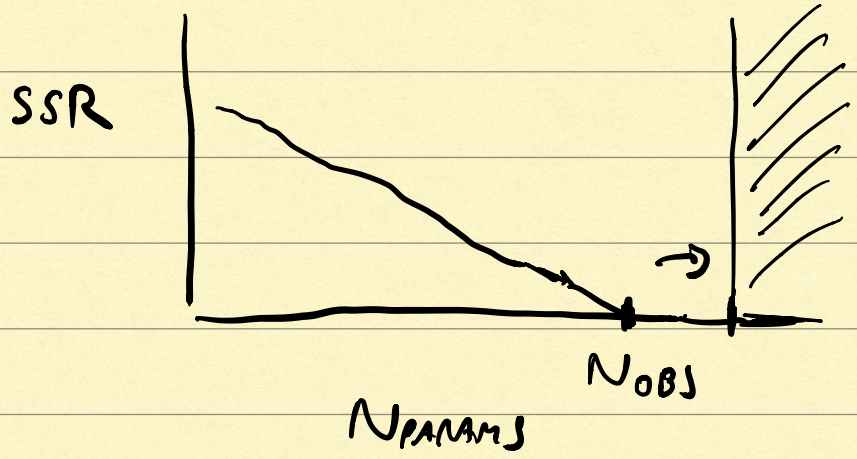
HIGH DIMENSIONAL REGRESSION

SUPPOSE N_{OBS} DATA POINTS

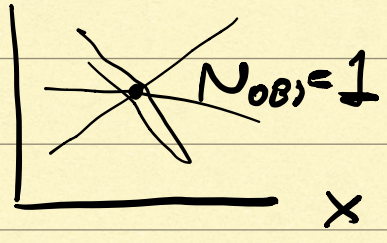
MODEL N_{PARAMS} PARAMETERS

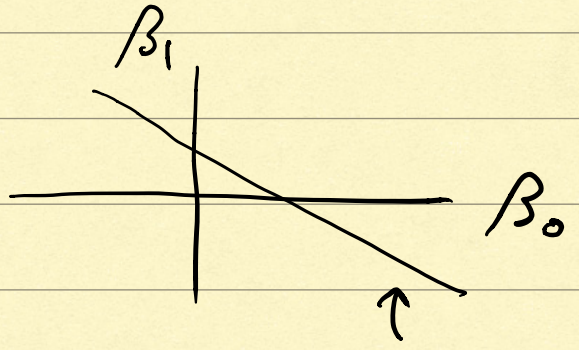


$N_{PARAMS} > N_{OBS}$

GENES	CELL
30,000	1000
BACTERIA IN GUT	PATIENTS
10^4	20

EX Y





$Y = \beta_0 + \beta_1 X$ $N_{PARAM}=2$

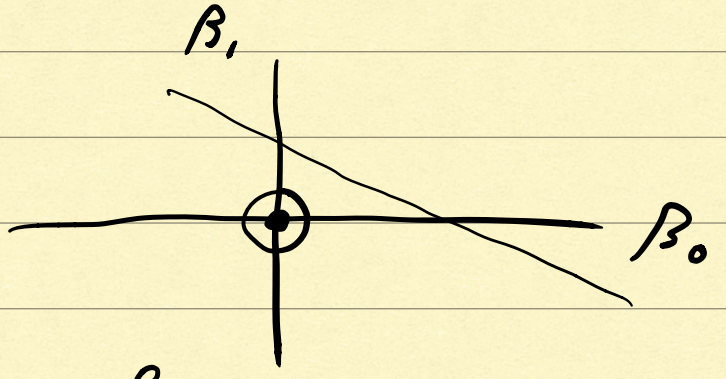
↑ ANY COMBO OF β_0, β_1 ON THIS CURVE

RIDGE REGRESSION (LAME)

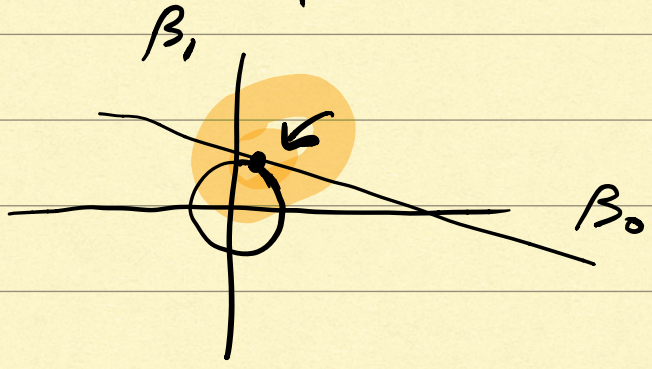
MINIMIZE $\log SSR + \lambda \sum_i (\beta_i)^2$

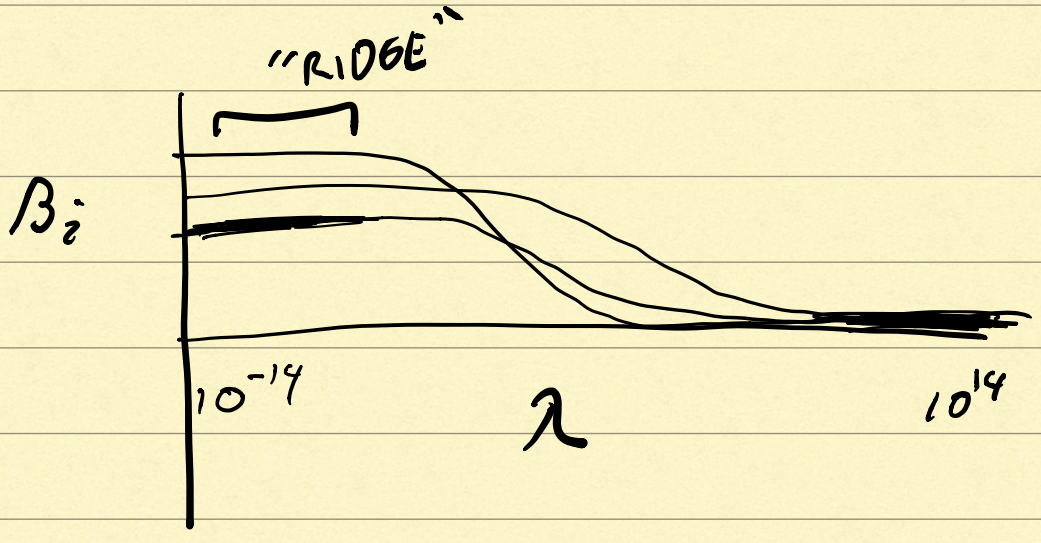
PENALTY

IF $\lambda \to \infty$



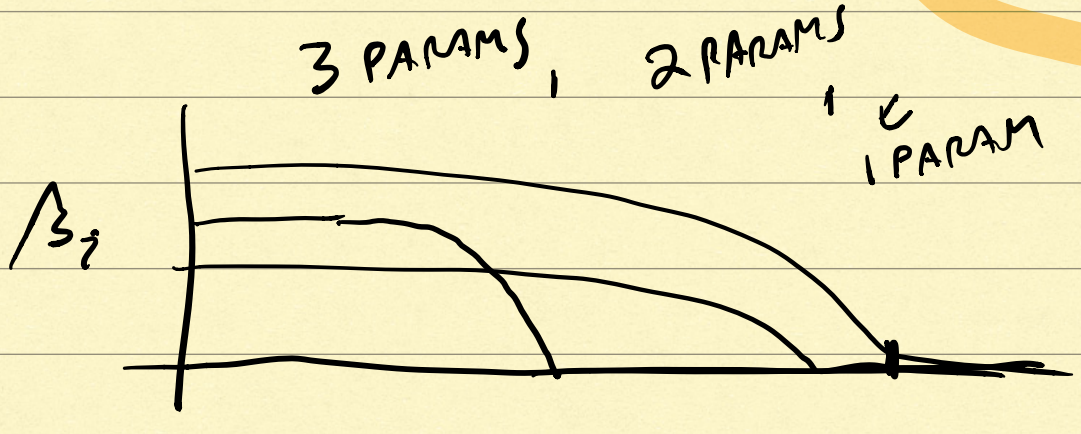
IF $\lambda \to 0$
(BUT NOT $\lambda = 0$)





LASSO REGRESSION

MINIMIZE $\log SSR + \lambda \sum_i |\beta_i|$

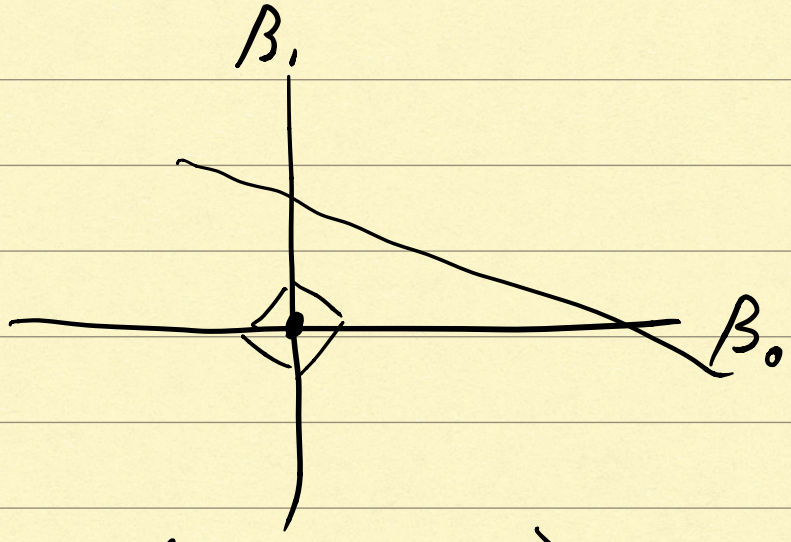


λ

← MODEL COMPLEXITY

$|\beta_1| + |\beta_2| =$ const

IF $\lambda \to \infty$



IF $\lambda \to 0$
(BUT NOT $\lambda = 0$)