$Y \quad X_1 \quad X_2$

LINEAR

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

PS 08

| ACTIVATION | SIZE | DEFORMABILITY |
|---|---|---|
| 0 | ⋮ | ⋮ |
| 1 | | |
| 0 | | |
| 1 | | |
| 1 | | |

df synth 1

TRAINING

GROUND TRUTH

$ ACTIVATION

Y TO LEARN

$ SIZE

$ DEFORMABILITY

X TO LEARN

df synth 2

VEIL OF SECRECY

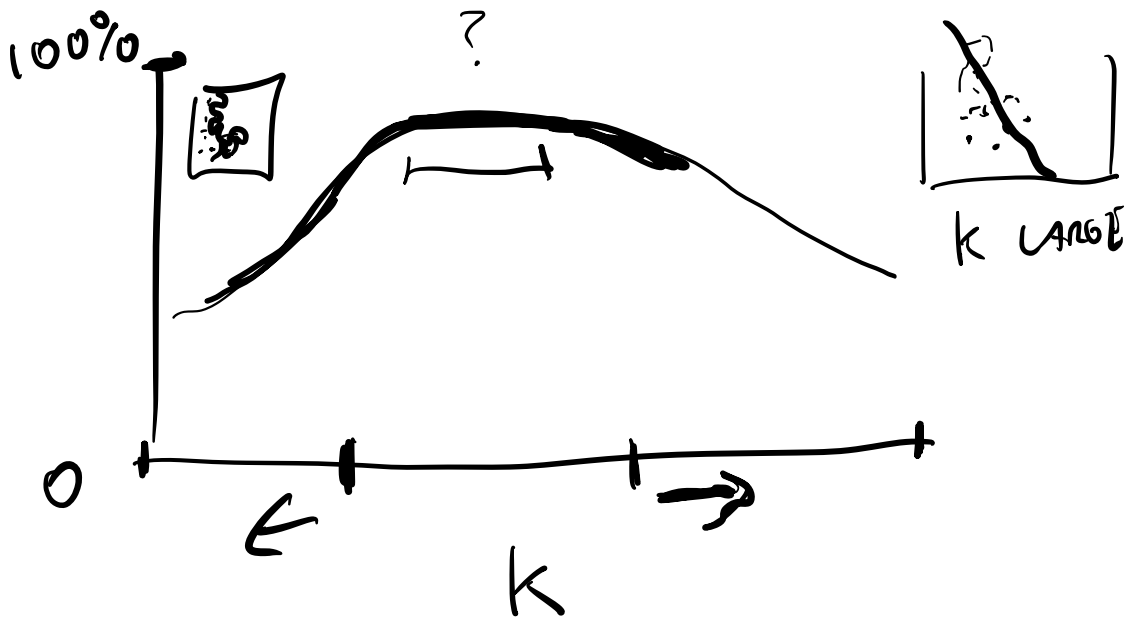
GROUND TRUTH TESTING

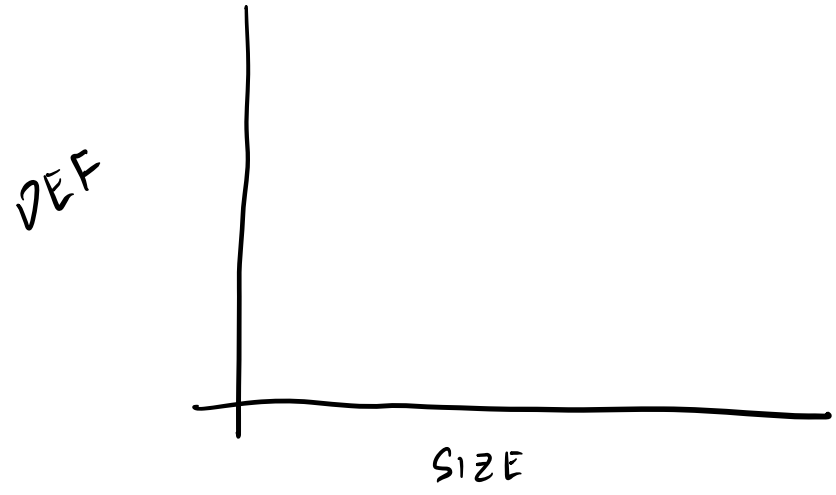
$ ACTIVATION
$ SIZE
$ DEFORM

X AT WHICH TO PREDICT

$ ACTIVATION ← $k_{NN}$ ( k )

PREDICTION TESTING

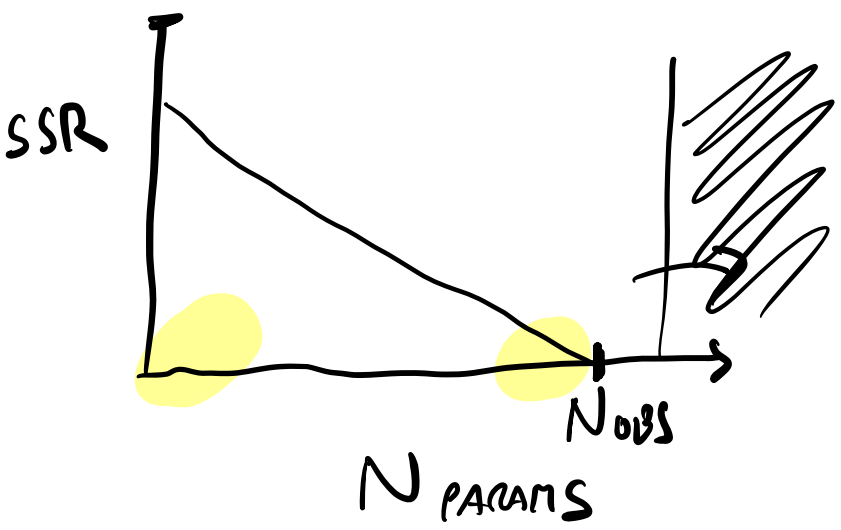





# HIGH DIMENSIONAL DATA

$N_{OBS}$ DATA POINTS

MODEL HAS $N_{PARAMS}$ PARAMETERS




$N_{PARAMS} > N_{OBS}$

NON IDENTIFIABLE

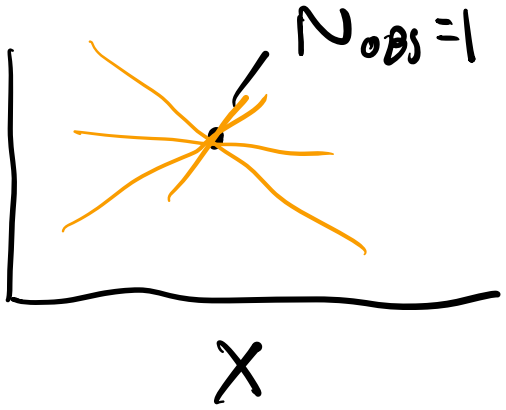
$N_{GENES} \sim 20000$

$N_{CELLS} \sim 1000s$ ?

BACTERIAL SPECIES IN GUT $10^4$

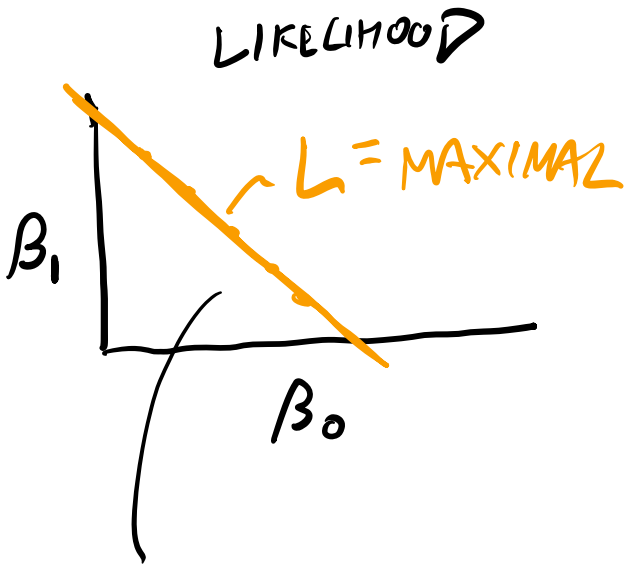
PATIENTS $\sim 10^?$

EX




$$Y = \beta_0 + \beta_1 X + \epsilon$$

$N_{PARAMS} = 2$




NO UNIQUE PARAMETERS THAT MAXIMIZE LIKELIHOOD

---

## RIDGE REGRESSION

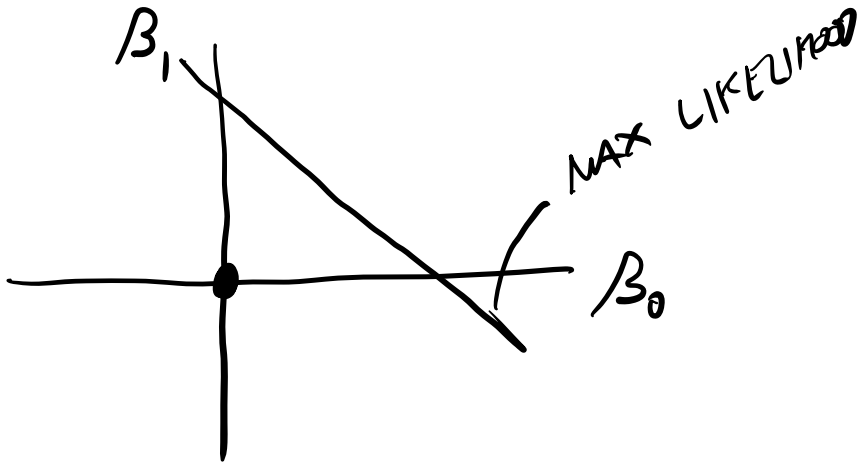
☹ LAME

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

EX $SSR + \lambda(\beta_0^2 + \beta_1^2)$

$\lambda$: CONSTANT

IF $\lambda \to \infty$




IF $\lambda \to 0$

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