2.

a) pop (all students in DS100) N $\longrightarrow$ sample n

small sample : doesn't matter

big sample : with replacement

b) our sample is fixed

$n=150 \rightarrow n=150_1 \rightarrow \hat{p}_1$

$n=150 \rightarrow n=150_2 \rightarrow \hat{p}_2$

$\vdots$

$n=150 \rightarrow n=150_{10000} \rightarrow \hat{p}_{10000}$

e.g. ABC $\rightarrow$ ABB

$\rightarrow$ CBB

w.r. $\rightarrow$ ACA

$\rightarrow$ ABC

ABC $\xrightarrow[\text{wor}]{}$ ABC

c) take $\hat{p}_1, \ldots \hat{p}_{10000}$ and look at middle 95% of values

left: 2.5 %ile
right: 97.5 %ile

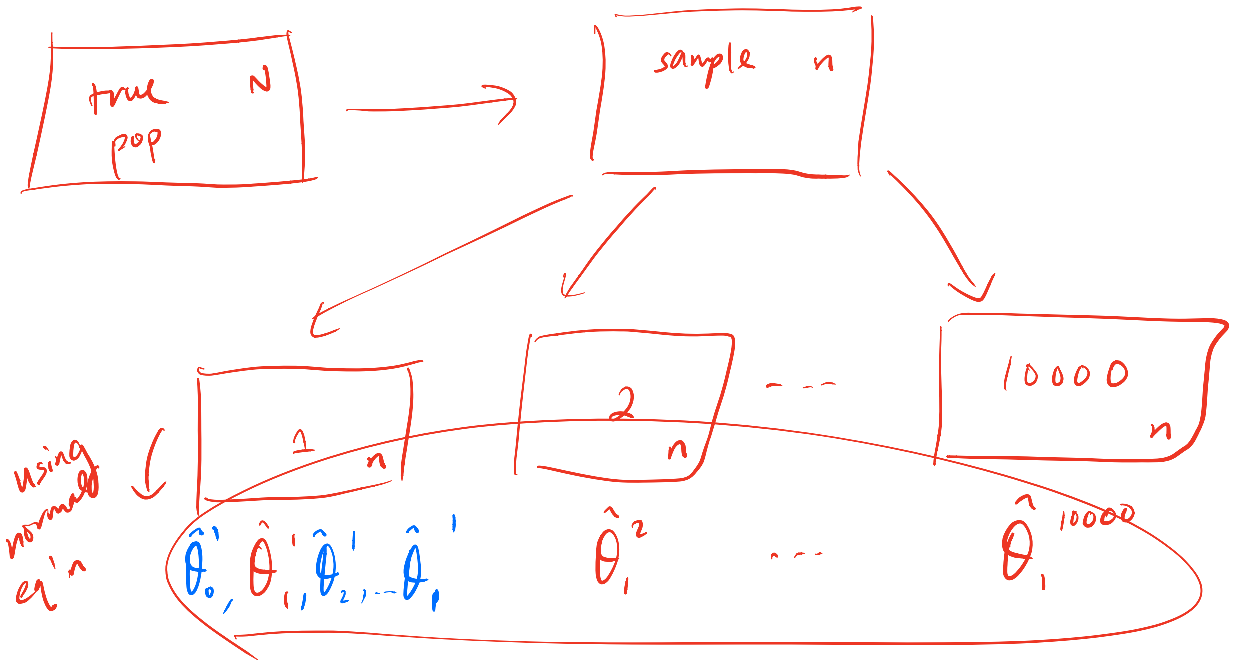
d) (2.3%, 4.7%)

$\boxed{\text{not}}$ : 95% chance that the true proportion is in this interval

true interpretation: if we repeat this process several times, 95% of the 95% CIs should contain the true p.

"this process": 1) collecting original sample
2) bootstrapping

3. $\mathbb{E}[y \mid x] = \theta_0 + \theta_1 x$




b) $f_\theta(x) = \theta_0 + \theta_1 x_1 + \dots + \theta_p x_p$


```
thetas = []
for i = 1, ..., 10000:
    bootstrap = sample with replacement (data, n)
    theta_hat = LinearModel.fit(bootstrap).getCoefficients
    thetas.append(theta_hat)
```

giving us $\hat{\theta}_0, \hat{\theta}_1, \dots \hat{\theta}_p$ for current resample

$$\text{thetas} = \begin{bmatrix} [\hat{\theta}_0^{1}, & \hat{\theta}_1^{1}, & \ldots & \hat{\theta}_p^{1}], \\ [\hat{\theta}_0^{2}, & \hat{\theta}_1^{2}, & \ldots & \hat{\theta}_p^{2}], \\ & \vdots & & \\ [\hat{\theta}_0^{10000}, & \hat{\theta}_1^{10000}, & \ldots & \hat{\theta}_p^{10000}] \end{bmatrix}$$

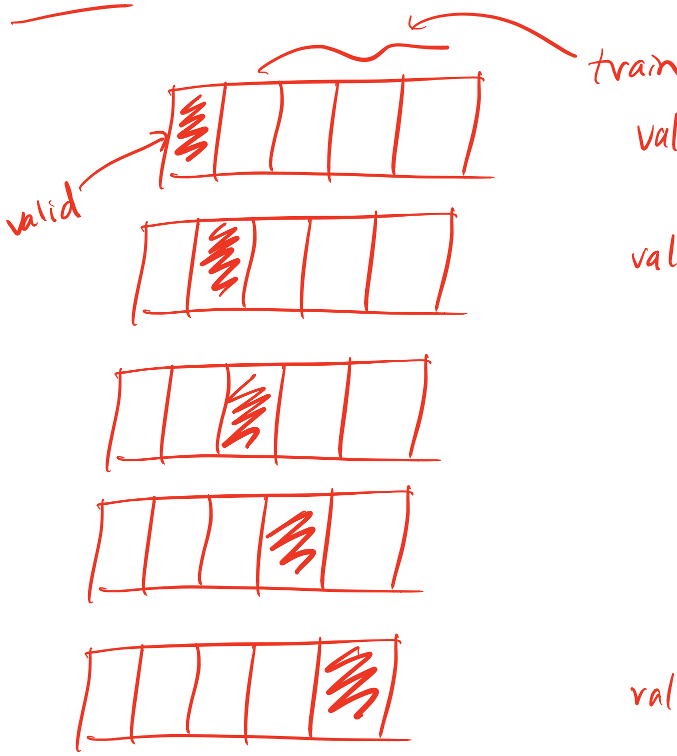
$$95\%\ CI = \begin{bmatrix} (2, 5) \\ (-3, 24) \\ \vdots \end{bmatrix}$$

1.

$$\theta^* = (X^T X)^{-1} X^T y$$

$$\theta^*_{ridge} = (X^T X + \lambda I)^{-1} X^T y$$

$\underline{k=5}$




valid err$_1$

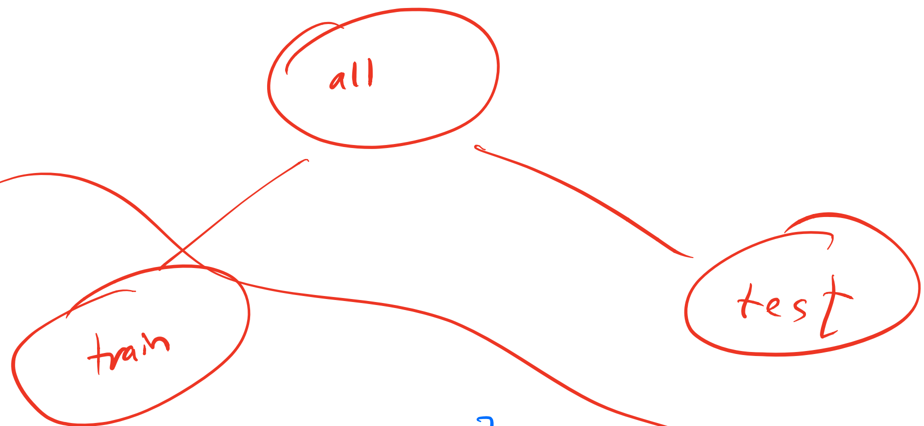
validerr$_2$

⋮

validerr$_5$

→ CV error

= mean validerr




num_trees = [1, 2, 5, 10, 20]

errors = [ ]

for each value ni in num_trees:

for value di in d:

compute CV error with depth = di

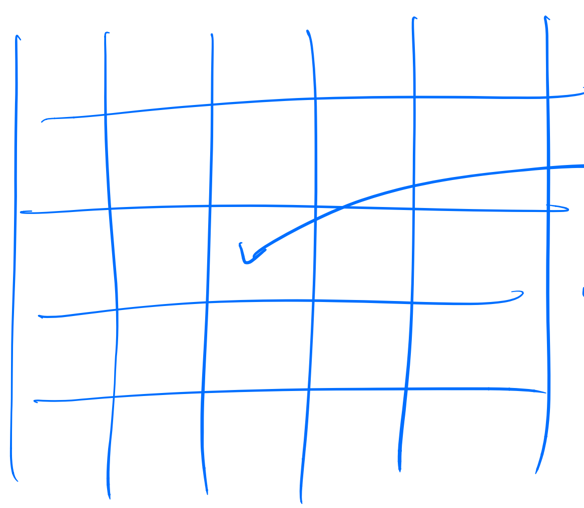
errors.append(error)

choose d with lowest error

~~errors~~

# trees

max depth



pick combo w/ min error