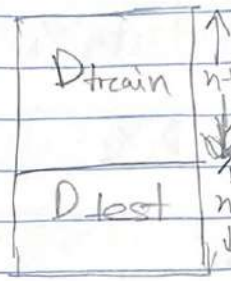
lecture 16



$$\frac{1}{k} = \frac{n_{test}}{n} \Rightarrow k = \frac{n}{n_{test}}$$

D
only a sample of the full distrib of $\langle X, Y \rangle$.

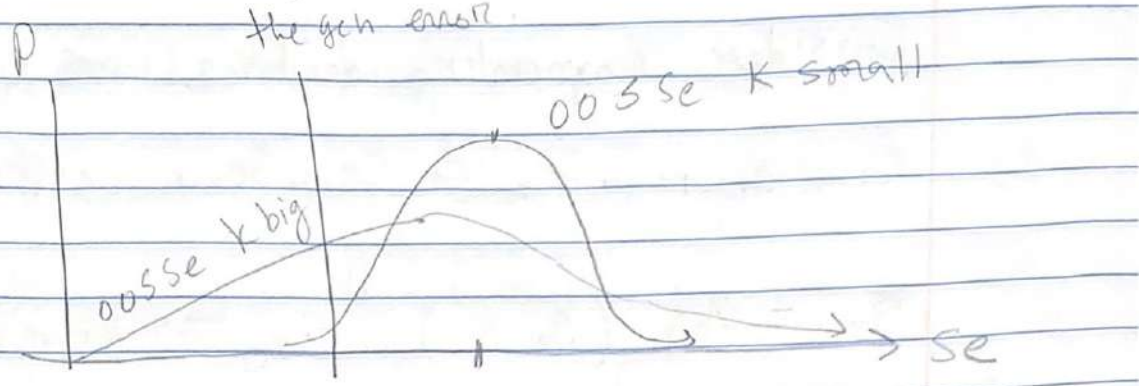
$$n_{test} \in \left\{1, 2, \ldots, \frac{n}{3}, \ldots \frac{n}{5}, \frac{n}{10}, n-1\right\}$$

$$\Rightarrow k \in \left\{n, \frac{n}{2}, \ldots, 2, 3, 5, 10, \frac{n}{n-1}\right\}$$

most common choices for k.

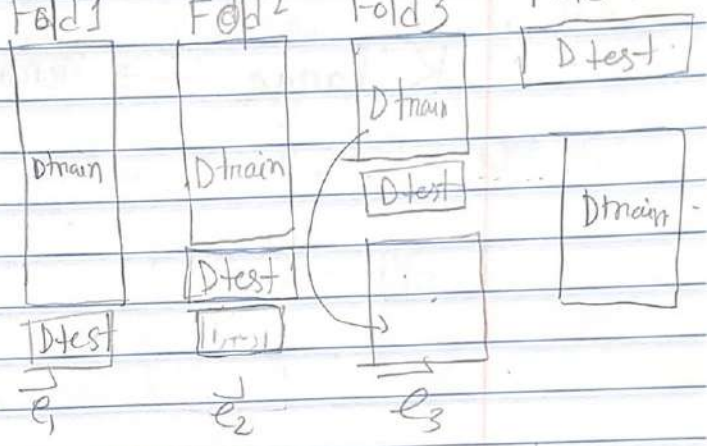
What's the trade off?

We want Validation; We use D_{test} to estimate "generalization error" AKA model performance. In the future. If n_{test} is small, the estimate is very variable but n_{test} is big... my expected estimate is closer to true performance of final



Cross Validation:- do many train-test splits.

K-Fold CV →



Another (later)

Overall K folds, each observation is oos once

If $\vec{e}_{CV} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$ has length n. Now Compute oos metrics on $\vec{e}_{CV}$.

If K = n (n folds) AKA "leave-one-out Cross Validation" (LOOCV)

Consider Computing metrics on

$\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_K \Rightarrow Se_1, Se_2, \ldots, Se_K$

$\overline{Se} = Avg\ (Se's), \quad S_{Se} = Stdev(Se's)$

$CI_{\sigma_e, 95\%} \approx [\overline{Se} \pm 2 S_{Se}]$

For this to be valid; $Se \overset{iid}{\sim} N(\ ,\ )$; Is this true? (Interpretable) NO.

K large $\rightarrow$ normality from CLT.

because e's are dependent since they are functions of g's which are dependent via sharing ID train g's

Computing a valid CI for generalization error under general condition is not possible.

Given D, You have many choices of model because there are many A's and H's. $g = \vec{A}(D, \vec{H})$ [change A and H, Therefore g change ✓]

Let's say you have a choice of a finite # of models, M: Choose between $g_1, g_2, \ldots g_m$? How do we Praw = 1. A: OLS

$Y = \mathbb{R}$.

$g_1 = b_0 + b_1 X$.

$g_2 = b_0 + b_1 X + b_2 X^2$

$g_3 = b_0 + b_1 \ln(X)$

$g_4 = b_0 + b_1 \mathbb{1}_{X \in [0,1]} + b_2 \mathbb{1}_{X \in [1,2]} + \ldots$

This is the fundamental problem of "Model selection."

In this class, we will provide an idea.
It is not the only idea

* g_1 is fit with A_1, H_1 on Dtrain and Se_1 is computed on ID_{test}.

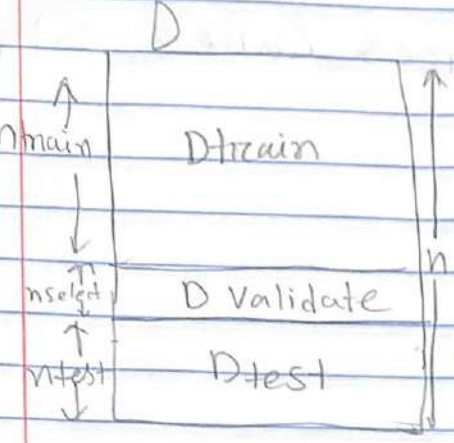
* g_2 is ~~with~~ fit with A_2, H_2 on Dtrain and Se_2 is computed D test

* g_m " " " A_m, H_m " " " Se_m "

Then select g_{m*} which has lowest Se.

Problems: (1) Se has high variance. So Pick K reasonable

(2) We have no longer honest Validation: Se is ~~not~~ honest because it could be overfit since ID_{test} was used many times.



Procedure

(1) Fit g_1 on D_{train}, Compute se_1 on D_{select}

or D_{select} (2) Fit g_2 on ..... se_2 ..

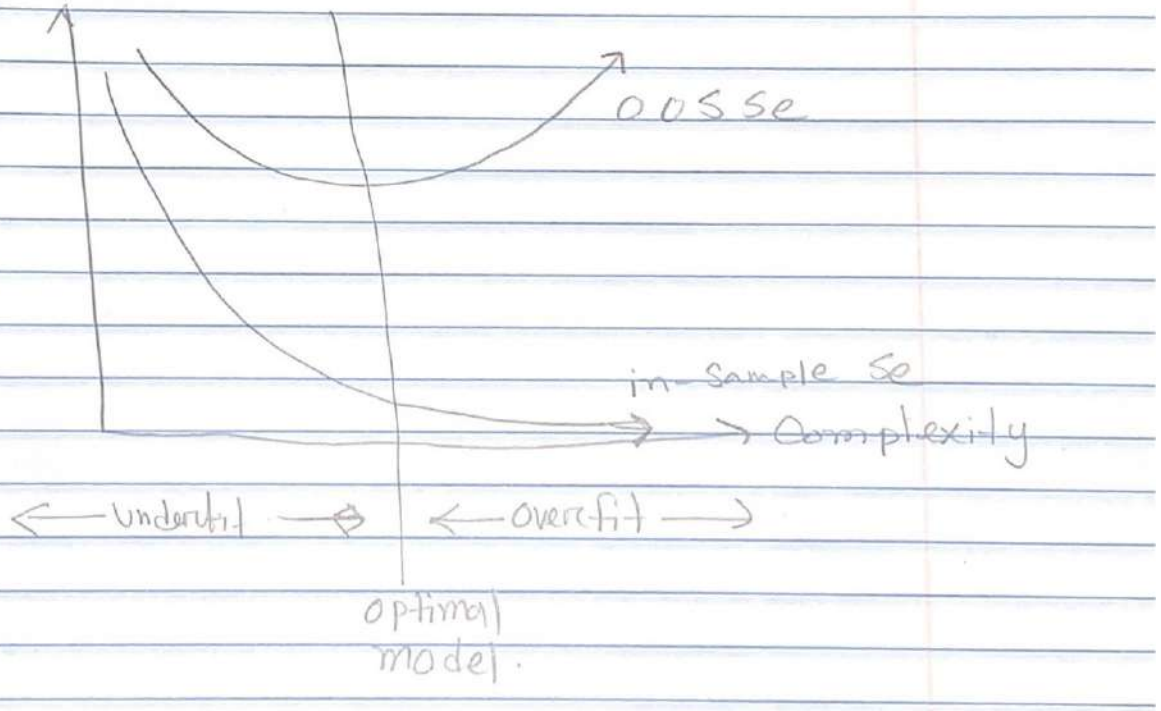
(m) Fit g_m on ... se_m ..

(m+1) final $m_x = \text{argmax}\{se_j\}$

(m+2) Fit g_{m_x} on $D_{train} \cup D_{select}$.

(m+3) Fit g_{final} on D.

uses of this procedure.



(1)

Consider $g_1, g_2, \ldots, g_m$ with increasing Complexity.

Complexity

This is Sometimes called Stepwise Modeling