Math 342W Lecture 15

$g(x) = a_1(x_1)^n + \dots + a_p(x_p)^n \, , \; n \geq 1$

↳ a_j's are continuous functions and this $g(x)$ is called a general additive model

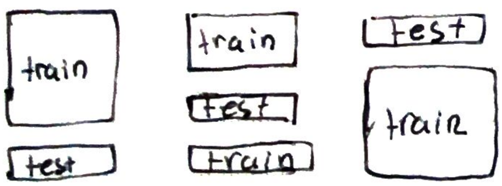
Interaction of features

By interacting features we are able to capture differential slopes of the given features.

→ Pretty hard to overfit with a large enough (n)

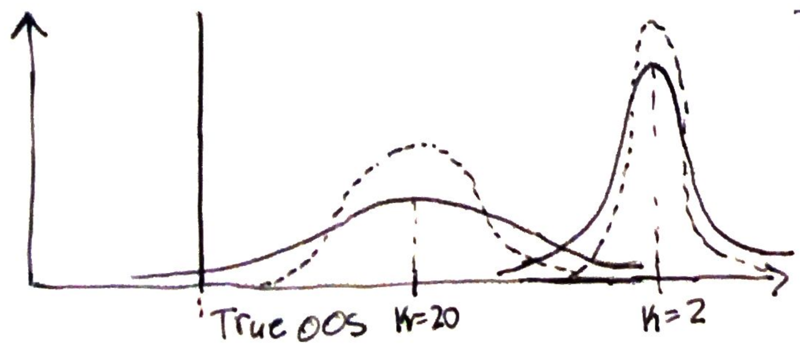
Validation / Cross Validation

The goal of these validations is to lower the variance in our metrics, particularly out of sample error metrics.



Each test from the folds returns $\vec{e}_i$ for the i^{th} fold

$$\vec{e}_{cv} = \begin{bmatrix} \vec{e}_1 \\ \vec{e}_2 \\ \vec{e}_3 \end{bmatrix} \Rightarrow oos\_SE = \sqrt{\frac{1}{n}\sum(\vec{e}_i - \bar{e})^2} \Rightarrow \begin{matrix} oos\_SE_1 \\ \vdots \\ oos\_SE_n \end{matrix} \Rightarrow \sqrt{\frac{1}{K-1}\sum(SE_\ell - \bar{SE})^2}$$



—— = Before C.V

- - - - = After C.V

⇒ Less Variance in $\underline{\underline{OOS}}$

(RMSE) @ K_i's

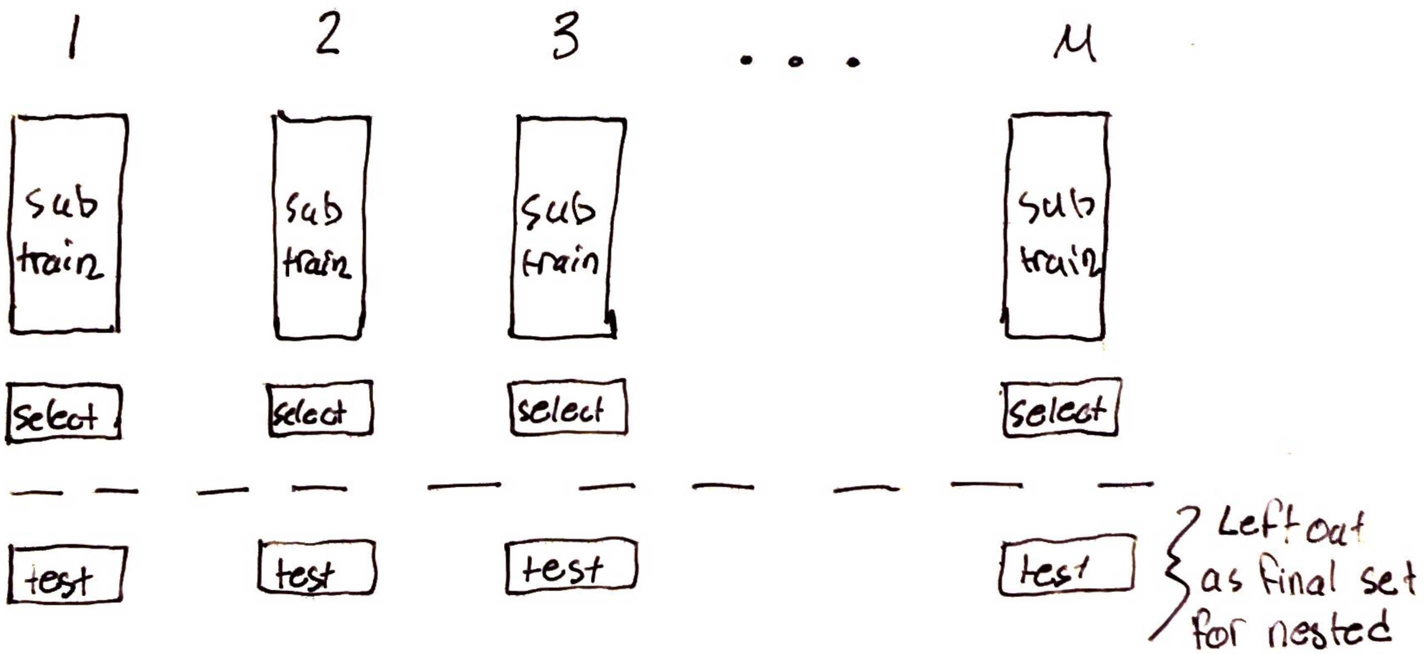
Model Selection

How do we pick the best of M models?

$g_1 = b_0 + b_1 x$
$g_2 = b_0 + b_1 x + b_2 x^2$
$g_3 = b_0 + b_1 \ln(x)$
$\vdots$
$g_m = $ ______

Our goal will be to select the model with the lowest out of sample error

We do a similar process to k-Fold CV in order to find the best model



1) We build each model (M) on $D_{subtrain}$
2) Evaluate model errors on D_{select} and pick best model m^*

→ The only issue now is that our OOS errors will be highly variable for m^*

More on how to fix this variability issue in Lec. 17 notes