2.) Goal: Explain why boosting using depth-one trees leads to an additive model of form: $f(X) = \sum_{j=1}^{p} f_j(X_j)$.

Let $b = 1, 2, \dots B$ index the trees.

Let $j = 1, 2, \dots p$ index the feature vars.

Let b_1 = feature var on which split is made in 1st tree; $b_1 \in [1, 2, \dots p]$

b_2 = " 2nd tree, etc.; $b_2 \in [1, 2, \dots p]$

etc.

Thus, we rewrite:

$$(8.12) \quad \hat{f}(x) = \sum_{b=1}^{B} \lambda \underbrace{\hat{f}^{b}(x)}_{\text{function of tree index}}$$

as...

function of feature var used in tree's sole split.

$$\circledast \quad \hat{f}(x) = \lambda \cdot \hat{f}(X_{b_1}) + \lambda \cdot \hat{f}(X_{b_2}) + \dots + \lambda \cdot \hat{f}(X_{b_B})$$

Now, if we let $b_{k,1}$ be 1st time kth feature var chosen, where $k \in [1, 2, \dots p]$

$b_{k,2}$ " 2nd " ,

we can rewrite $\circledast$ as: (by collecting all instances that kth var is chosen for split into a bracket) for $k \in [1, 2, \dots p]$

$$\hat{f}(x) = [\lambda \cdot \hat{f}(X_{b_{11}}) + \lambda \cdot \hat{f}(X_{b_{12}}) + \dots] + [\lambda \cdot \hat{f}(X_{b_{21}}) + \lambda \cdot \hat{f}(X_{b_{22}}) + \dots] + \dots + \underbrace{[\lambda \cdot \hat{f}(X_{b_{p1}}) + \lambda \cdot \hat{f}(X_{b_{p2}}) + \dots]}_{= f_p(X_p)}$$

Finally, letting: $[\lambda \cdot \hat{f}(X_{b_{j1}}) + \lambda \hat{f}(X_{b_{j2}}) + \dots] = f_j(X_j)$ for $\forall j$

we get:

$$f(X) = f_1(X_1) + \dots + f_p(X_p) = \boxed{\sum_{j=1}^{p} f_j(X_j)}$$