Note: These are **my** notations that we are using today (and that will be used in the afternoon assignment, which is why I've chosen to use them as well).

Generic terms (e.g. they don't have the subscript m, and aren't specific to a stage in the learning process)

$G(X)$: An ensemble of boosted weak learners (today, trees).

α: A generic weighting factor (also referred to as the learning rate/shrinkage parameter), used to determine the size of the update in the sequential learning process.

γ: A set of hyper-parameters (for our decision trees, this could be the max depth, minimum samples per leaf, etc.).

X : Our full matrix of inputs.

$\phi(X, \gamma)$: An unlearned model/machine learning algorithm (today, a decision tree) that takes X as input and has hyper-parameters γ_m.

$L(y_i, \phi(x_i, \gamma))$: The calculated loss between the target value y_i and the output of our weak learner with x_i as input and γ as the hyper-parameters to optimize over.

Terms of a specific stage/iteration in the learning process (e.g. they have a subscript m)

$G_m(X)$: Stage m weak learner that has been trained (today, trees).

α_m: A learned weighting factor (learning rate/shrinkage parameter), used to update the sequential learning process in dynamic ways (e.g. potentially used to weight updates that perform better more heavily). In practice, with boosted trees we just have a fixed learning rate.

γ_m: A specific of hyper-parameters that has been learned (optimized over) at stage m (so, maybe we chose a max depth of 6, a minimum number of samples per leaf of 10, etc.).

r_{im} : The gradient of the loss function with respect to some stage $m-1$ weak learner, for observation i.
