Algorithm. *Super Learning for Sequential Prediction in Longitudinal Data*

For each rule d:

⋆ Obtain an estimator $Q^d_{T+1,n}$ of $Q^d_{T+1}(L_{0:T})$ with super learning.

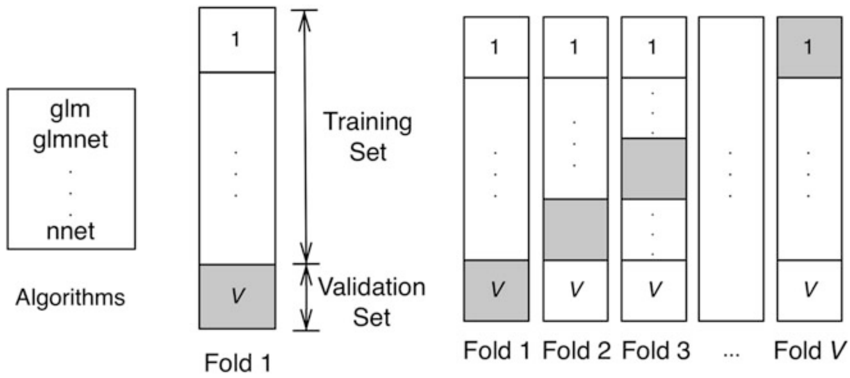
For $t = T + 1$ to $t = 1$

⋆ Define $Q^d_{t,n}(L_{0:t-1})$ as the outcome in next regression and use super learning to estimate $E(Q^d_t(L_{0:t-1}) \mid A_{t-2:0} = d(L_{t-2:0}), L_{t-2:0})$.

Save the final estimator $Q^d_{t=1,n}(L_0)$ as estimator of $\bar{Q}_{Y^d} = E(Y^d \mid L_0)$.

Specifically, the sequential super learner for $\bar{Q}_{Y^d}$ is constructed as follows:

1. Let $Q^d_{T+2,n} = Y$.
2. Set $t = T + 1$.
3. For time point t, create a data set of n observations where each observation has an outcome $Q^d_{t+1,n}(L_{t+1:0})$, and covariates $A_{0:t}, L_{0:t}$. Fit the K candidate regression algorithms within V-fold cross-validation. Recall that $B_n \in \{0, 1\}^n$ is a random variable that splits the data into a training set $\{i : B_n(i) = 0\}$ and validation set $\{i : B_n(i) = 1\}$. The data set is divided into a training set containing $\frac{V-1}{V}^{\text{ths}}$ of the data and a validation set containing the remaining $\frac{1}{V}^{\text{th}}$of the data in each of V folds. For each $v = 1, \ldots, V$, for each $k = 1, \ldots, K$, train the k-th algorithm on the training set $T(v)$, while the $V(v)$ validation set is run through the fitted algorithm to obtain cross-validated predicted values. This results in a predicted value $Z^d_{k,t,i}$ for each algorithm k and subject i, $i = 1, \ldots, n$.



4. Posit a family of weighted combinations of the K algorithms that is a convex combination indexed by α, and select the α_n that minimizes the cross-validated empirical mean of the loss function.