

1. Split data
into V blocks

2. Train each
candidate learner

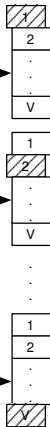
3. Predict the outcomes in the
validation block based on the
corresponding training block
candidate learner

4. Model selection and
fitting for the regression
of the observed
outcome onto the
predicted outcomes
from the candidate
learners

5. Evaluate super learner
by combining predictions from
each candidate learner (step 0)
with $m(z; \beta)$ (steps 1-4)

0. Train each
candidate learner on
entire dataset

Data



lm D/S/A ... RF

lm D/S/A ... RF

lm D/S/A ... RF

lm D/S/A ... RF

Z				obs.
lm	D/S/A	...	RF	Y
1	1		1	1
2	2		2	2
.	.		.	.
.	.		.	.
.	.		.	.
V	V		V	V

$$E(Y | Z) = m(z; \beta)$$

Super Learner