

DATA 202 — Ridge and Lasso · Week 10 Wednesday

Same houses as Monday: 2,193 training houses in Ames, Iowa, and 5-fold cross-validation. Your professor will tell you when to do each part.

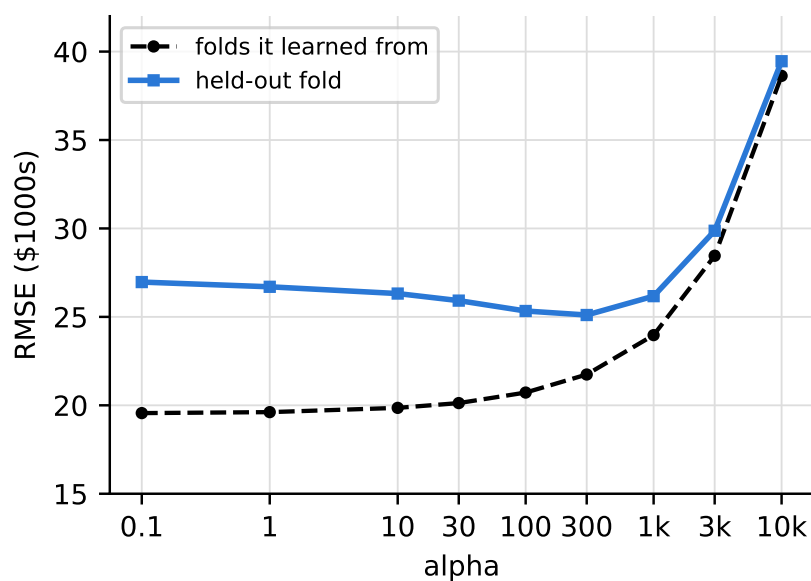
1 · More freedom

RMSE (typical miss, in dollars) of plain least squares, with and without the product columns.

model	columns	folds it learned from	held-out fold	gap
Monday	47	\$27,069	\$27,876	_____
with products	237	\$19,548	\$28,958	_____

- Which model is overfitting? _____ How can you tell? _____
- The products made the misses on *known* houses _____ and on *new* houses _____.

2 · Choose alpha



- Ridge's typical miss for each alpha.
- Circle** the best alpha.
- Label** the overfitting side and the underfitting side.
- Here a *larger* knob means a _____ model. For kNN, a *larger* k meant a _____ model.

3 · What Lasso keeps

alpha	columns kept (of 237)	CV RMSE
10	212	\$26,365
30	195	\$25,775
100	155	\$24,644
300	105	\$24,222
1,000	58	\$24,510
3,000	31	\$26,079
10,000	18	\$31,393

1. **Circle** the alpha with the lowest CV RMSE.
 2. Find the **simplest** model within \$500 of the best. alpha = _____, with _____ columns.
 3. Lasso dropped `second_floor_area` but kept `living_area`. Why might that cost almost nothing?

-