

S.1

- ② For prediction, you have a few, noisy variables $\rightarrow$ ~~Ridge~~ Linear regression
- ④ " , you have a substantial number of noise variables $\rightarrow$ LASSO
- ③ For explanation, you have about 80 variables $\rightarrow$ LASSO
- ② " , you have 10 variables and all are important $\rightarrow$ ~~Logistic~~ Ridge regression

S.2

$$\text{LASSO} = \frac{\text{RSS}}{n} + \lambda \sum_{j=1}^p |B_j|$$

As λ increases, what happens? = Means the model is getting simpler and coefficients shrink

- ④ The MSE on the test data initially decreases and eventually increase again

Simpler model $\rightarrow$ Variance $\downarrow \rightarrow$ model becomes more stable and less sensitive to training data,

improve test performance for a bit. But bias starts to $\uparrow$ and

the model becomes too simple and miss real patterns

- ⑥ The MSE on training data steadily increases

- ③ The irreducible error on test data remains constant (cannot be affected by the model)

- ④ The squared bias on the test data steadily increases (simpler model $\rightarrow$ more bias)

- ② The variance on test data steadily decreases