



Fictional example for the model

$y = 0.2X_1 + 0.2X_2 + 0.2X_3 + 0.2X_4 + 0.2X_5 + \epsilon$  of 100 observations,  
 $\epsilon \sim \mathcal{N}(0, 1)$ .  $X_1$ - $X_4$  are independently drawn from different normal  
distributions:  $X_1, X_2, X_3, X_4 \sim \mathcal{N}(0, 2)$ . While  $X_1$ - $X_4$  have pairwise  
correlation coefficients of 0,  $X_4$  and  $X_5$  are nearly perfectly correlated:  
 $X_5 = X_4 + \delta, \delta \sim \mathcal{N}(0, 0.3), \rho(X_4, X_5) = 0.98$ .

We see that Lasso shrinks the coefficient for  $X_5$  to zero early on, while  
Ridge assigns similar coefficients to  $X_4, X_5$  for larger  $\lambda$ .