Exercise 1:

Consider the following data generating process with $n = 100$ observations and p covariates. Initially set the number of predictors $p = 3$ and $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$. $\mathbf{\Sigma}$ is the covariance matrix with the variance on the diagonal and small values on the off-diagonal (both values chosen by you). The true coefficients are initially given as $\beta = (0.5, 0.5, -0.5)$ and the errors are drawn from a normal distribution $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 = 10)$.

a) Write a function to calculate the ridge regression estimator in closed form (depending on λ) as discussed in the lecture.

b) Specify a sequence grid of lambdas containing 100 values and calculate the ridge regression coefficients using the function above. Plot the coefficient paths (λ is on the x-axis) for all three coefficients for one draw.

c) Generate a test sample of the same size and calculate the prediction error and the training error for each value of λ and for the OLS estimate. Show your results graphically.

d) Replace one of the covariates with a constant equal to -10. Show that including the constant (without de-centering first) affects the estimated coefficient paths as discussed in the lecture.

e) Using the `glmnet` package, determine the optimal lambda from a 10-fold cross-validation procedure.

Exercise 2 (Simulation Study)**:**

a) Compare the performance of the OLS estimator and the ridge regression estimator in terms of the prediction error for 100 simulation runs

b) Propose a manipulation of the dgp that will improve the ridge regression performance relative to the OLS estimate.