DSP2025 Homework 5 Due date: 12:00 noon, May. 12, 2025

1. A linear time-invariant filter is described by the difference equation

$$y[n] = -0.8y[n-1] + 0.8x[n] + x[n-1]$$

(a) Determine the system function $H(z)$ for this system. Express $H(z)$ as a ratio of polynomials in z^{-1}(negative powers of z) and also as a ratio of polynomials in positive powers of z.

(b) Plot the poles and zeros of $H(z)$ in the z-plane.

(c) From $H(z)$ obtain an expression for $H(e^{j\hat{w}})$, the frequency response of this system. Show that $\left|H\left(e^{j\hat{w}}\right)\right|^2 = 1$ for all $\hat{w}$.

2. Write a small program to perform the following task:

Using the given dataset from https://datarepository.wolframcloud.com/resources/Sample-Data-Birth-Weight-Risk, which contains 189 observations of newborns, implement and compare Elastic Net and Group Lasso to predict birth weight (BWT) from eight predictors.

Your task is to write a single Jupyter notebook that performs the following steps:

1. Data loading
2. Feature construction
   (a) Create third-order polynomial features for Age and LWT.

Polynomial Features (first 5):

	Age	LWT	Age^2	Age LWT	LWT^2	Age^3	Age^2 LWT	Age LWT^2	LWT^3
0	19.0	182.0	361.0	3458.0	33124.0	6859.0	65702.0	629356.0	6028568.0
1	33.0	155.0	1089.0	5115.0	24025.0	35937.0	168795.0	792825.0	3723875.0
2	20.0	105.0	400.0	2100.0	11025.0	8000.0	42000.0	220500.0	1157625.0
3	21.0	108.0	441.0	2268.0	11664.0	9261.0	47628.0	244944.0	1259712.0
4	18.0	107.0	324.0	1926.0	11449.0	5832.0	34668.0	206082.0	1225043.0

(b) One-hot encode all six categorical predictors, retaining every dummy column.

One-hot Encoded Features (first 5):

	Race_1	Race_2	Race_3	Smoker_False	Smoker_True	PTL_0	PTL_1	PTL_2	PTL_3	Hypertension_False	Hypertension_True
0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0
1	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0
2	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0
3	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0
4	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0

3. Standardization
   (a) Standardize the full feature matrix (zero mean, unit variance).
4. Elastic Net
   (a) Use ElasticNetCV to tune alpha and l1_ratio with 3-fold Cross Validation (CV).
   (b) Report the best (alpha, l1_ratio), the 3-fold CV mean squared error (MSE), and list all features with non-zero coefficients.
5. Group Lasso
   Import and use the Group Lasso implementation from the group_lasso package
   (a) Define feature groups
   {Group 0: Age
   Group 1: LWT
   Group 2: Age*LWT
   Group 3: Race
   Group 4: Smoker
   Group 5: PTL
   Group 6: Hypertension
   Group 7: UI
   Group 8: FTV}
   (b) Use a 3-fold CV to tune the group_reg and l1_reg.
   (c) Report the best (group_reg, l1_reg), its CV MSE, and the list of groups with non-zero coefficients.
6. Results summary
   Displays the summary table of Elastic Net and Group Lasso

Reference:

[1] Hosmer, D. W. and Lemeshow, S. (1989) Applied Logistic Regression. New York: Wiley

[2] Yuan, Ming, and Yi Lin. "Model selection and estimation in regression with grouped variables." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 68.1 (2006): 49-67.