1.

In a linear regression problem with $n = 15$ data points, the sample variance of the dependent variable is 1000. We use forward selection to choose the best model with $d \in \{1, 2, \ldots, 8\}$ parameters. The regression sum of squares RegSS of the model selected for each d is shown in the following table.

d	RegSS
1	0
2	1011
3	1513
4	1826
5	2885
6	3674
7	4519
8	5672

Assume that we want to construct a regularized logistic regression model for this dataset.

(a) Which model is going to be selected by Mallow's Cp (AIC) and Adjusted R^2?

(b) Test at $\alpha = 0.01$ the overall significance of the model selected by Adjusted R^2.

Note: contrary to what you might think, you do not need extra information to solve this problem. Everything you need to solve this problem is in this exam.

2.

Assume that we have a Ridge regression problem with only one predictor, and the true model is linear *without an intercept*, i.e. $Y = \beta_1 X + \epsilon$. Assume that we have n samples, $(x_i, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ and we want to find the $\mathscr{L}_2$ regularized least squares estimate $\hat{\beta}_1$ from the data.

(a) Formulate the objective function in terms of a candidate $\hat{\beta}_1$ and x_i's and y_i's, which are known. Assume that the regularization parameter is λ.

(b) Find $\hat{\beta}_1$ in terms of λ and the data,.

3.

For the following data set for classification:

Index	X	Y
1	-1	1
2	0	0
3	3	0

Assume that we want to construct a regularized logistic regression model for this dataset.

(a) Write down the $\mathscr{L}_2$-regularized loss function $J(\beta_0, \beta_1)$ for this dataset with regularization parameter $\lambda = 1$.

(b) Compare the bias variance of the regularized model with the unregularized model ($\lambda = 0$).

4.

For the following data set for classification:

Index	X	Y
1	-1	1
2	0	0
3	3	0

Assume that we want to construct a regularized logistic regression model for this dataset.

(a) Write down the $\mathscr{L}_2$-regularized loss function $J(\beta_0, \beta_1)$ for this dataset with regularization parameter $\lambda = 1$.

(b) Compare the bias variance of the regularized model with the unregularized model ($\lambda = 0$).

5.

Assume that in a linear regression problem, we have four features X_1, X_2, X_3, X_4 and $X_3 = 4.45X_1 - 6.87$. Explain why each of the following terms are valid or invalid shrinkage penalties in general. For those that are valid shrinkage penalties, why they are appropriate or inappropriate for this particular problem with four features X_1, X_2, X_3, X_4 and $X_3 = 4.45X_1 - 6.87$:

(a) β_2^2.

(b) $\beta_1^5 + \beta_2^5 + \beta_3^5 + \beta_4^5$.

(c) $|\beta_1| + \beta_2^2 + |\beta_3| + \beta_4^6$.

(d) $\beta_1^2 + |\beta_2| + \beta_3^6 + |\beta_4|$.

(e) $\sqrt{|\beta_1|} + \beta_3^2$

6.

Consider the following dataset:

Index	X_1	X_2	Y
1	0	-1	1
2	1	0	2
3	0	1	3
4	1	1	2

We wish to fit a regularized multinomial regression model to this dataset. Write down the $\mathscr{L}_1$-penalized negative log likelihood function for this dataset. Use one parameter vector $(\beta_{0k}, \beta_{1k}, \ldots, \beta_{pk}), k = 1, 2, 3$ for each class.

7.

Assume that using some data, we find the least squares solution for a linear regression problem involving two predictors is $(\beta_0, \beta_1, \beta_2) = (1.25, 6.36, 8.53)$.

(a) Assume that we regularize the problem by adding the Ridge constraint $\beta_1^2 + \beta_2^2 \leq 16$. Does the regularizer shrink (change) the least squares solution? Explain.

(b) Assume we regularize the problem by adding the LASSO constraint $|\beta_1| + |\beta_2| \leq 25$. Does the regularizer shrink (change) the least squares solution? Explain.

Hint: In this problem, we used constraints, NOT regularizer penalties. Also, plotting the constraints in β_1 - β_2 plane may help you understand the problem.