Q1) a) ① Lasso regression stands for Least Absolute Shrinkage and Selection Operator

② It is a type of linear regression that includes a regularization term.

③ The regularization term is L1 penalty.

④ The goal of lasso regression is to minimum the sum of RSS i.e. Residual sum of squares with the absolute values of coefficients by tuning parameter $\lambda$

$$\text{minimize} \left( \sum_{i=1}^{n} (y_i - \hat{y_i})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right)$$

where,

$y_i$ = observed response
$\hat{y_i}$ = predicted response,
$\beta_j$ = model coefficients
$\lambda$ = regularization parameter.

⑤ Feature Selection in Lasso :-
Lasso performs automatic feature section by forcing coefficients of some features to be exactly zero.
The process is as :-
a) L1 penalty added to cost function encourages sparsity.

b) feature with coefficients shrinking to zero are removed model

c) This allows lasso to select only the most important features, making the model more efficient and reducing the multicolinearity.

Q1) b) ① Linear Regression :-
Models the relationship between a dependent variable and independent variable using a straight line.

② Ridge Regression :-
A variant of linear regression with L2 regularization to prevent overfitting

③ Lasso Regression :-
Similar to Ridge but uses L1 regularization, which shrink some zero and performing feature selection.

④ Elastic Net Regression:
Combine L1 and L2 by balancing feature selection.

(5) Polynomial Regression:
Extends linear regression by modelling non-linear relationships using polynomial terms.

(6) Logistic Regression:
Used for binary classification, models and outcomes of 0 or 1.

(7) KNN :-
K-Nearest Neighbors Regression. It predicts the value based on average of k nearest neighbors.

Q1)(c) (1) Bias Variance Trade off

(1) Bias :-
The error is by or comes by assuming a model that don't capture the underlying models of data.
high bias leads to overfitting

(2) Variance :-
The error is due to models sensitivity to the data. High variance leads to overfitting, where the model becomes too complex.

* Relationship to Underfitting & Overfitting

(1) Underfitting occurs when the model has high bias and poor performance

(2) Overfitting occurs when the model has high variance and gives good performance.

(3) a) (1) Kernel methods in SVM are used to transform non-linearly data into a higher-dimensional space where it becomes linearly separable.

(2) This is achieved without computing the coordinates of data in this higher-dimensional space.

* Common Kernel functions

(1) Linear kernel :-
No transformation is applied, and the data is assumed to be linearly separated.

(2) Polynomial Kernel :-
Maps the data into a higher-dimensional polynomial feature