What are the types of Regression Techniques in Supervised Learning?

Regression Analysis:

Regression analysis is a predictive analytical method that analyzes the relationship between a target or a variable and independent variable that differs from the database. Different types of regression analysis techniques are used where targeted and independent variables show direct or indirect linear relationships between each other, and target differences than continuous values. The regression process is mainly used to determine predictability, weather patterns, time series, and in case of cause and effect.

Regression analysis is a key way to solve setback problems in machine learning using data modeling. It involves determining the appropriate line, which is the line that passes through all the data points in such a way that the distance of the line from each data point is reduced.

Types of Regression Techniques in Supervised Learning

1. Linear Regression
2. Logistic Regression
3. Ridge Regression
4. Lasso Regression

Linear Regression:

Linear Regression is one of the most basic types of machine deviation in machine learning. The line order model consists of prediction variables and dependent variables related to each other. In the event that data involves more than one independent variable, Linear Regression is termed by Multiple Regression order models.

The equation given below is used to show the Linear Regression model:

$y = mx + c + e$

Where m is the line slope, c breaks(intercepts), and e represents an error in the model.

Logistic regression:

This type of regression analysis is used when the dependent variable is binary in nature. For example, if the outcome of interest is death in a cancer study, any patient in the study can have only one of two possible outcomes- dead or alive. The impact of one or more predictor variables on this binary variable is assessed. The predictor variables can be either quantitative or qualitative. Unlike linear regression, this type of regression does not require a linear relationship between the predictor and dependent variables. For logistic regression to be meaningful, the following criteria must be met/satisfied.

The equation given below is used to show the Linear Regression model:
y = e^(b0 + b1*x) / (1 + e^(b0 + b1*x))

Where y is the predicted output, b0 is the bias or intercept term and b1 is the coefficient for the single input value (x). Each column in your input data has an associated b coefficient (a constant real value) that must be learned from your training data.

Ridge Regression:

Ridge Regression is a technique for analyzing multiple regression data that suffer from multicollinearity. When multicollinearity occurs, least squares estimates are unbiased, but their variances are large so they may be far from the true value. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors.

The equation given below is used to show the Ridge Regression model:

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{m} \hat{\beta}_j^2 = ||y - X\hat{\beta}||^2 + \lambda||\hat{\beta}||^2.$$

where The λ parameter is the regularization penalty.

Ridge Regression, which penalizes sum of squared coefficients (L2 penalty).
LASSO Regression:

LASSO Regression is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model.

Residual Sum of Squares + λ * (Sum of the absolute value of the magnitude of coefficients)

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Where, λ denotes the amount of shrinkage.

Lasso Regression, which penalizes the sum of absolute values of the coefficients (L1 penalty).

Applications of Regression Analysis:
There are three major uses of regression analysis – attributing causality, forecasting and prediction.

Implementation:
Implementation on diabetes dataset(regression dataset) different regressor as follows,

Load dataset and important libraries:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.datasets import load_diabetes
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score, mean_squared_error
from sklearn.linear_model import LinearRegression, Ridge, Lasso
```

Split dataset for training and testing purpose:
Load regression dataset diabetes using sklearn.datasets and split for training and testing purpose,

```
x,y = load_diabetes(return_X_y=True)
x = x[:, np.newaxis, 2]
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.
20,random_state=32)
```

Create the model, train the model and predict for testing data:
Create different models for different regressors as follows,

```
model1=LinearRegression()
model1.fit(x_train,y_train)
y_pred_1=model1.predict(x_test)
model2=Ridge()
model2.fit(x_train,y_train)
y_pred_2=model2.predict(x_test)
model3=Lasso()
model3.fit(x_train,y_train)
y_pred_3=model3.predict(x_test)
```

Evaluation:
Generate the R2 score of each model as follows,

```
r2_score_1 = model1.score(x_test,y_test)
print("Linear Regression Score",r2_score_1*100,'%')
r2_score_2 = model2.score(x_test,y_test)
print("Ridge Score",r2_score_2*100,'%')
r2_score_3 = model3.score(x_test,y_test)
print("Lasso Score",r2_score_3*100,'%')
```

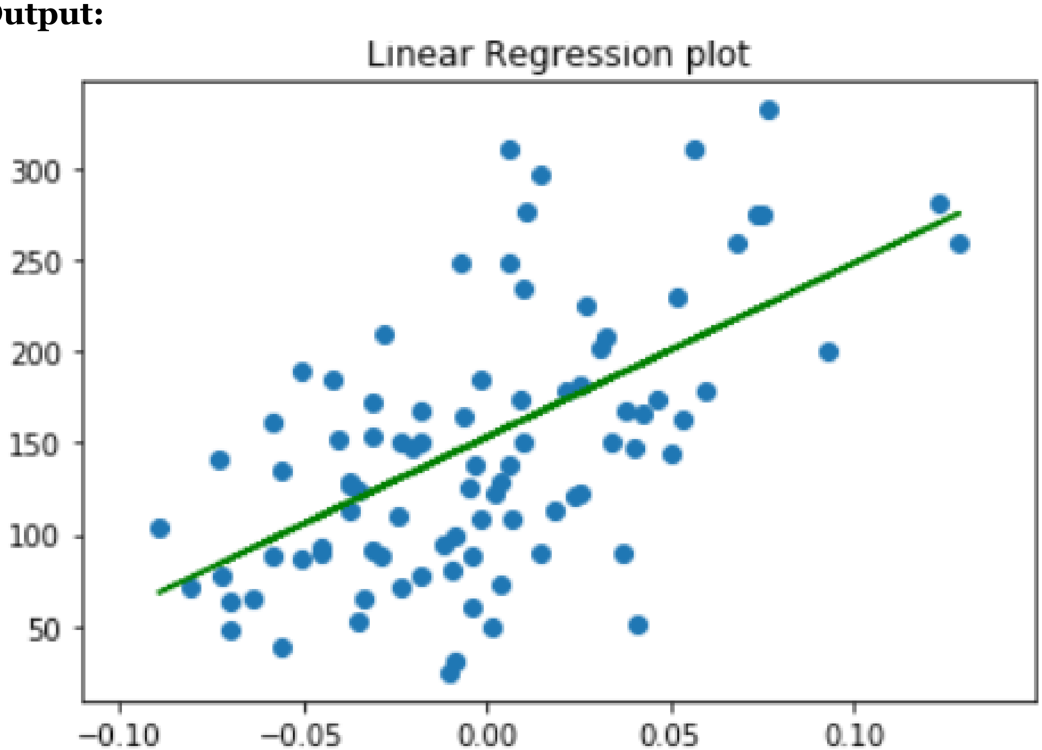
Output:
More is the value better is the algorithm performance,

```
Linear Regression Score 45.222834232361954 %
Ridge Score 40.878119097154055 %
Lasso Score 36.468710812077674 %
```

Regression model plot:
`Linear Regression plot`

```
import matplotlib.pyplot as plt
plt.scatter(x_test,y_test)
plt.title("Linear Regression plot")
plt.plot(x_test, y_pred_1, color='green')
plt.show()
```

Output:



Lasso Regression plot

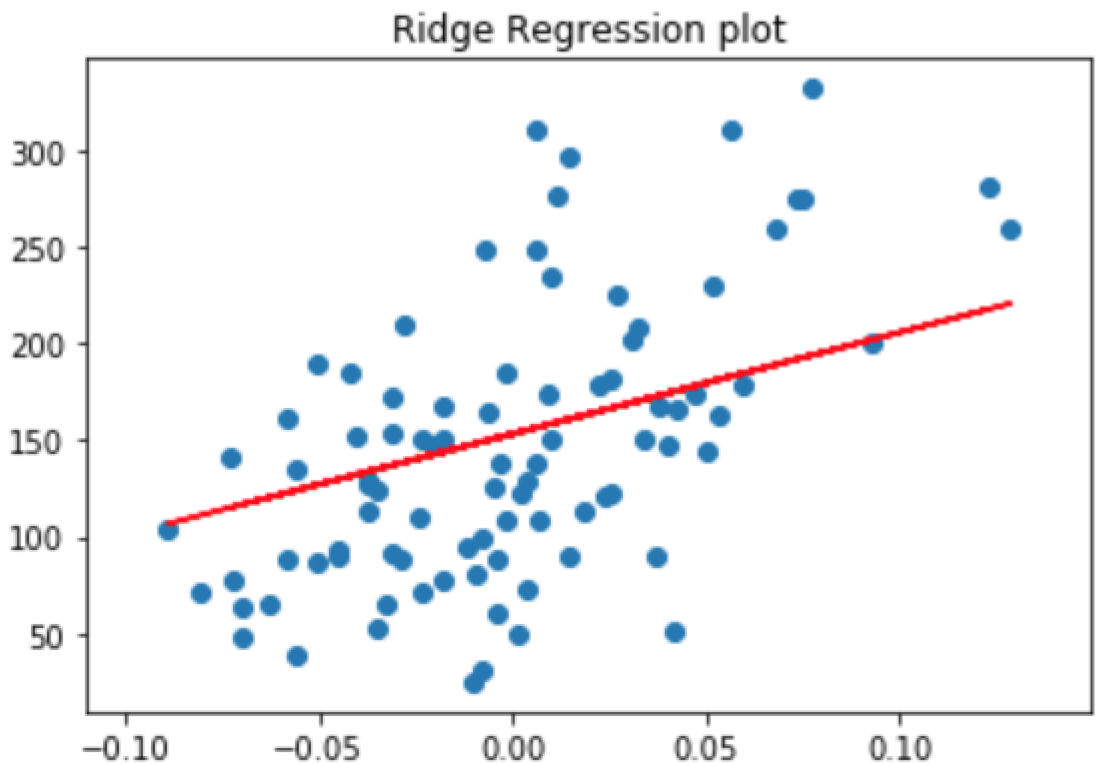
```
plt.scatter(x_test,y_test)
plt.title("Lasso Regression plot")
plt.plot(x_test, y_pred_2, color='k')
plt.show()
```

Output:

Ridge Regression plot

```
plt.scatter(x_test,y_test)
plt.title("Ridge Regression plot")
plt.plot(x_test, y_pred_3, color='red')
plt.show()
```

Output:



Conclude:

From visualization and scores it is confirmed that linear regression algorithm performance is good.