Documentation



MACHINE LEARNING



STL-10 (image Data set) && Titanic (numerical Data set)

A)GENERAL INFORMATION ON DATA SET

1-numerical data set

name	Titanic data set
Total number of samples(rows)	1309
features(inputs)	11(PassengerId,Pclass,Name,Sex,Age,SibSp,Parch,Ticket,Fare,Cabin,Embark ed)
target variable (predict)	1(supervised)
training sample	80%
validation &testing sample	20%

* preprocessing steps

filling missing values in columns

```
( Age and Fare using the median)
(Embarked (char) using mode)
```

Creating new Feature for better learning performance

These features aim to capture potentially useful information about passengers' social connections aboard the Titanic.

- FimalySize`Feature
 - `SibSp`Number of siblings and spouses aboard.
 - `Parch` Numer of Parents and Childern aboard.
 - adding one for the passanger themselves
 - making FamilySize a comprehensive measure of a passenger's group size.
- `ISAlone` Feature

This feature identifies passengers traveling alone

- Encoding char columns using one Hot Encoding in Label Encoder
 - -drop unused columns due to serveral reasons
 - Columns are `Cabin`, `SibSp`, `Parch`
 - High Missing Rate `Cabin` Feature
 - low predictice power
 - to reduce the noise and simplify the dataset

* Details about used modules

1-KNeighborsClassifier

Parameter

```
`n_neighbors=5`:
```

- Indicates the number of nearest neighbors to consider when making a prediction.
- For classification, the label of a new data point is determined by majority voting among its 5 closest neighbors in the feature space.

2- LogisticRegression

Parameters

```
`max_iter=1000`:
```

- Specifies the maximum number of iterations for the solver to converge.
- -Logistic regression uses an optimization algorithm (e.g., gradient descent or one of the solvers like lbfgs) to find the best-fit coefficients.
- This parameter ensures the solver doesn't run indefinitely if it struggles to converge.

- Default value is 100. Increasing it (e.g., to 1000) can help the model converge when the dataset is complex or poorly scaled.

2-IMAGE data set

name	STL-10
classes and labels	Air plane , bird , car , cat ,deer , dog ,horse , monkey , ship ,truck
Total image	6000
image size	32*32
training image	5000 image
	80%
Validation & testing sample	1000 image 20%

***** Over view about data set

- ➤ 10 classes: horse, monkey, cat, dog, car, ship, bird, truck, deer, air plane
- ➤ Images are 96*96 pixels ,color
- > 500 training images per class (10 predefined flods), 800test image per class
- ➤ 100000 unlabeled image for unsupervised learning, these example are extracted from similar but broader distribution of images

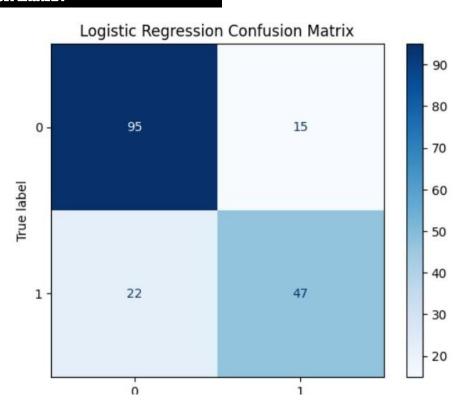
➤ Image were aquire from labeled example on image net .

- &preprocessing steps
 - -resize
 - -normalization
 - -convert to tensors /array
 - -feature extraction (cnn)

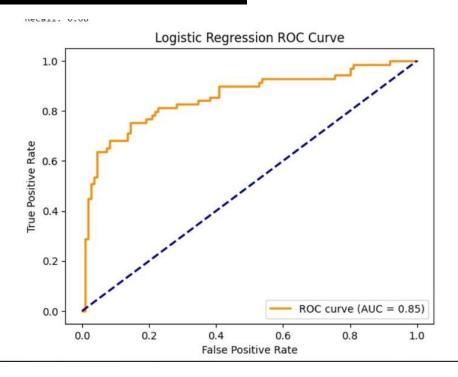
B) IMPELEMENTATION DETAILS

1-LOGISTIC REGRATION as a linear

accuracy	0.79
Loss value	0.45
Confusion matrix	



Precision	0.76
recall	0.68



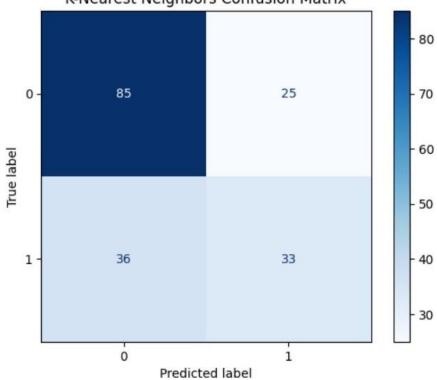
	0.85
auc graph	0.00

2-KNN REGRESSOR

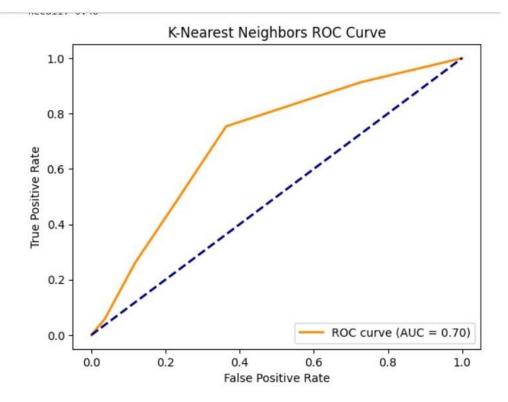
accuracy	0.66
Loss value	0.34

Confusion matrix

K-Nearest Neighbors Confusion Matrix



Precision	0.57
recall	0.48

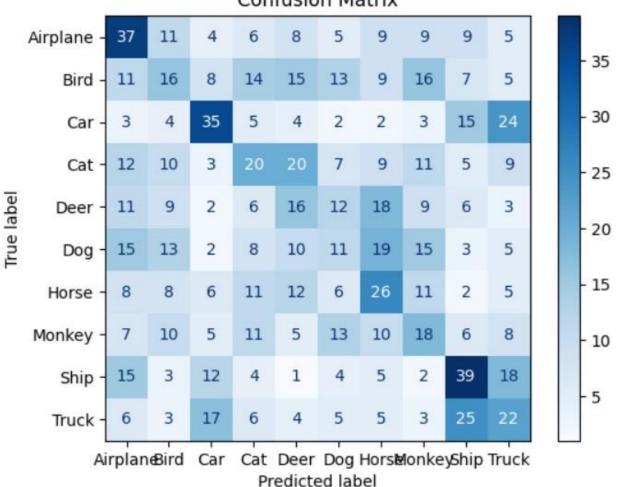


auc graph 0.70

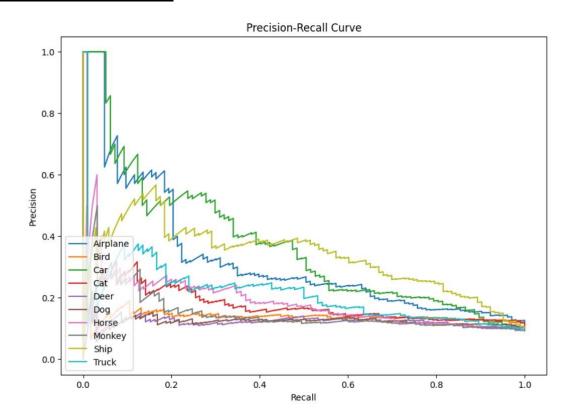
3-LOGISTIC REGRATION

accuracy	0.24
Loss value	2.788
Confusion matrix	

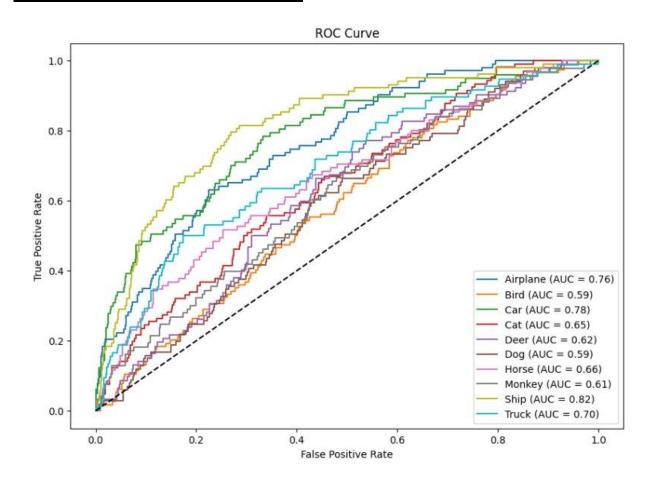
Confusion Matrix

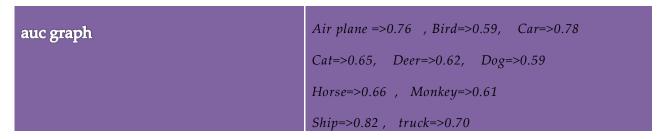


Precision and recall



Classification	n Report:			
	precision	recall	f1-score	support
Airplane	0.30	0.36	0.32	103
Bird	0.18	0.14	0.16	114
Car	0.37	0.36	0.37	97
Cat	0.22	0.19	0.20	106
Deer	0.17	0.17	0.17	92
Dog	0.14	0.11	0.12	101
Horse	0.23	0.27	0.25	95
Monkey	0.19	0.19	0.19	93
Ship	0.33	0.38	0.35	103
Truck	0.21	0.23	0.22	96
accuracy			0.24	1000
macro avg	0.23	0.24	0.24	1000
weighted avg	0.23	0.24	0.24	1000

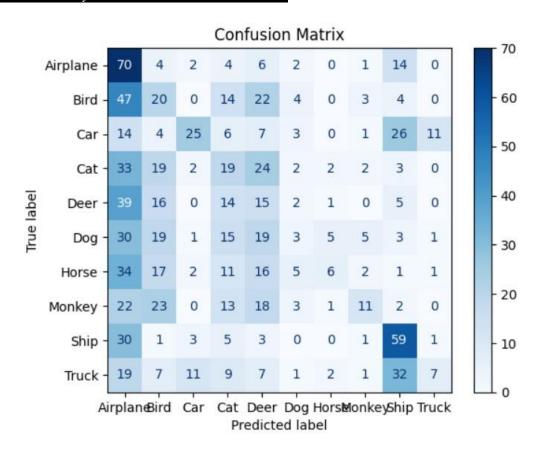




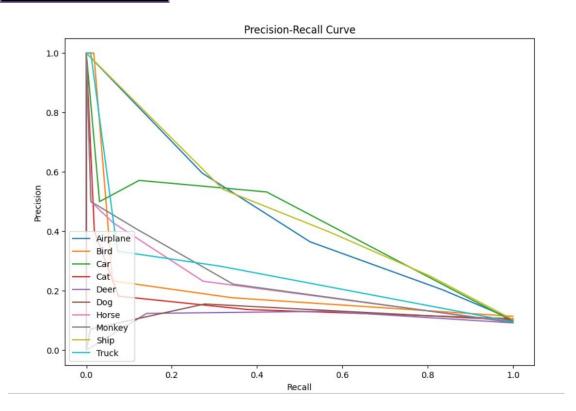
4-KNN CLASSIFIER

accuracy	0.23
Loss value	0.765

Confusion matrix

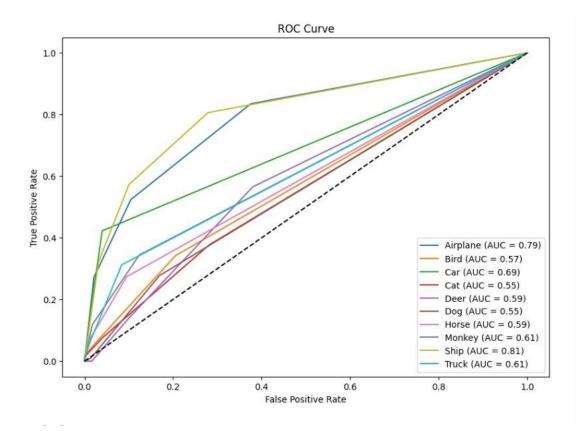


Precision and recall



Classification Report:	Class	ificat	ion Re	port:
------------------------	-------	--------	--------	-------

	precision	recall	f1-score	support
Airplane	0.21	0.68	0.32	103
Bird	0.15	0.18	0.16	114
Car	0.54	0.26	0.35	97
Cat	0.17	0.18	0.18	106
Deer	0.11	0.16	0.13	92
Dog	0.12	0.03	0.05	101
Horse	0.35	0.06	0.11	95
Monkey	0.41	0.12	0.18	93
Ship	0.40	0.57	0.47	103
Truck	0.33	0.07	0.12	96
accuracy			0.23	1000
macro avg	0.28	0.23	0.21	1000
weighted avg	0.28	0.23	0.21	1000



Auc graph

Air plane =>0.79 , Bird=>0.57, Car=>0.69

Cat=>0.55, Deer=>0.59, Dog=>0.55

Horse=>0.59 , Monkey=>0.61

Ship=>0.81 , truck=>0.61