

پروژه اول – بررسی تاثیر تقسیم بندی داده ها به مجموعه های آموزش و تست در دقت الگوریتم

بیان مساله

در این پروژه داده های ما مربوط به یک باغ وحش است و هدف الگوریتم دسته بندی حیوانات است.

در مجموعه داده ('Zoo.csv') اسم حیوانات و مشخصه های هر کدام نظیر پرداختن، تخم گذار بودن، تعداد پاهای دم داشتن و غیره به همراه شماره دسته آورده شده است. در مجموعه داده ('class.csv') دسته های پستاندار، ماهیها، حشرات و... به همراه شماره دسته آمده است.

مرحله صفر: اضافه کردن کتابخانه های مورد نیاز

```
import numpy as np # linear algebra
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import os
from sklearn.neural_network import MLPClassifier
from sklearn import preprocessing
%matplotlib inline
```

کتابخانه های مربوط به محاسبات عددی، رسم نمودارها، دریافت مجموعه داده ها و خواندن آنها و تابع مربوط به شبکه عصبی پرسپترون چند لایه به پروژه اضافه می شود.

مرحله یک: خواندن مجموعه داده ها

```
animal = pd.read_csv('zoo.csv')
animal.head(10)
```

مرحله دوم: پیش پردازش داده ها و مشاهده نمودار توزیع

در مرحله پیش پردازش، ابتدا با کلید قرار دادن شماره دسته ها (Class_Number) دو مجموعه داده را به هم متصل میکنیم.

```
ani_class = pd.read_csv('class.csv')
df = pd.merge(animal,ani_class,how='left',left_on='class_type',right_on='Class_Number')
df.head()
```

؟ با استفاده از تابع factorplot از کتابخانه Seaborn نمودار مربوط به تعداد حیوانات را با دسته بندی موجود در مجموعه مجموعه داده ('class.csv') رسم کنید.

؟ آیا این مجموعه داده، مجموعه متعادل است؟ در مورد مجموعه های نامتعادل و روش های ایجاد تعادل در مجموعه داده های نامتعادل تحقیق کنید و بررسی کنید در صورتیکه این مجموعه داده نامتعادل است چگونه میتوان آن را متعادل کرد و چه تاثیری در عملکرد مدل دارد؟

مرحله سوم: تقسیم بندی داده ها

؟ تقسیم بندی داده ها را با استفاده از تابع `train_test_split` از کتابخانه `sklearn.model_selection` را با توجه به تعداد داده های موجود در دو دسته آموزش و آزمون تقسیم کنید.

مرحله چهارم: آموزش شبکه عصبی پرسپترونی چند لایه

؟ سپس مدل شبکه عصبی پرسپترون چند لایه را آموزش داده و دقت آن را بررسی کنید. از تابع `MLPClassifier` استفاده کنید.

مرحله پنجم: بحث و بررسی نتایج

؟ مرحله بالا را با مقادیر دیگر آزمایش کنید و تغییرات عملکرد الگوریتم را بررسی کنید.