

به نام خدا

گزارش سوالات برنامه نویسی تمرین سری سوم درس یادگیری ماشین

زهرا دهقانیان

سوال ۷)

a) ابتدا در تابع `load_data` از کتابخانه `sklearn` مجموعه داده `wine` را `load` کردیم. سپس برچسب داده ها و ۱۳ ویژگی دیگر را در کنار هم در یک ماتریس قرار دادیم و بازگردانیدیم.

سپس تابع `evaluate_algorithm` که در تمرین قبلی برای `K-fold cross validation` استفاده کردیم را، فراخواندیم. در این تابع داده ها را به ۶ بخش تقسیم کردیم و ۶ مرتبه الگوریتم `naïve bayes` را با هر یک از این مجموعه ها به عنوان داده تست و بقیه به عنوان داده آموزش فراخواندیم. در تابع `naïve_bayes` ابتدا تابع `summarize_by_class` را روی مجموعه داده آموزش فراخواندیم. در این تابع، ابتدا داده ها آموزش بر حسب کلاس داده ها (۱ و ۲ و ۳) از یکدیگر جدا می شوند و سپس برای هر کدام از کلاس ها تابع `summarize_dataset` (که وظیفه محاسبه میانگین و واریانس و تعداد اعضای هر کلاس را دارد)، را صدا میزنیم. سپس تابع `predict` را با اطلاعات کلاس ها و داده های تست فرامیخوانیم. در تابع `predict`، تابع `calculate_class_probability` را فرامیخوانیم. در این تابع برای هر کلاس، به ازای هر کدام از ۱۳ ویژگی، مقدار `Gaussian_probability` محاسبه میشود و در هم ضرب می شود و به عنوان احتمال آن کلاس برگردانده میشود. و سپس کلاسی که بیشترین احتمال را دارد، به عنوان کلاس تخمین زده شده، برگردانده میشود. در نهایت با مقایسه برچسب های تخمین زده شده و برچسب های اصلی دقت الگوریتم براس هر دسته و دقت میانگین محاسبه میشود.

دقت الگوریتم برای هر دسته و دقت میانگین به صورت زیر است :

```

C:\Users\zdehg\AppData\Local\Programs\Python\Python37\python.exe
D:/univesity/fogelisans/machine_learning/HW3/Q7/Cross_Validation.py
C:\Users\zdehg\AppData\Local\Programs\Python\Python37\lib\site-
packages\sklearn\externals\joblib\externals\cloudpickle\cloudpickle.py
:47: DeprecationWarning: the imp module is deprecated in favour of
importlib; see the module's documentation for alternative uses
import imp
Scores: [96.55172413793103, 100.0, 100.0, 100.0, 100.0, 100.0]
Mean Accuracy: 99.425%

Process finished with exit code 0

```

(b) برای این بخش از پیاده سازی بخش قبل استفاده میکنیم و تنها به جای k بار اجرا الگوریتم، ۱ بار و با تقسیم داده به نسبت ۲ به ۱ برای داده آموزش و تست اجرا میکنیم. برای رسم نمودار ROC با توجه به این که مسئله اصلی، ۳ کلاس دارد، از روش one Vs. all استفاده میکنیم و پس از تخمین برچسب ها تابع one_vs_all را فرا میخوانیم تا هر یک از کلاس ها را یک بار class_of_interest قرار دهد و بقیه را منفی بگیرد و در نهایت معیار ارزیابی ROC را از کتابخانه sklearn.metrics بارگذاری میکنیم و نمودار را به ازای هر کدام از کلاس ها میکشیم. نمودارهای ROC برای کلاس ها به صورت زیر است :

