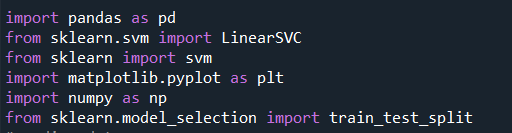
داده کاوی

تمرین سری سوم

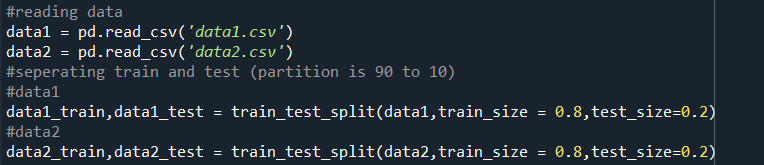
تمرین عملی

سوال 1)

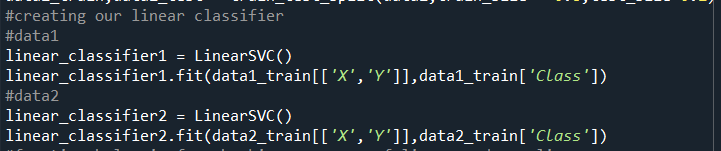
اول از همه کتاب خانه های مورد نیاز را import میکنیم.



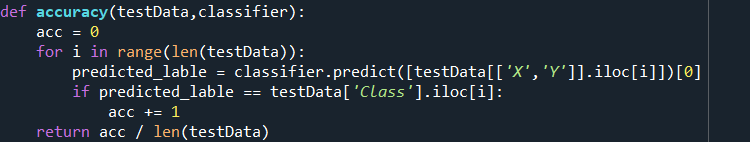
حال داده ها را میخوانیم و train و test را به نسبت 4 به 1(0.8 داده ها اموزش و 2/0 داده ها برای تست)



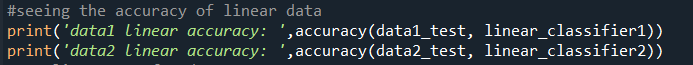
حال دسته بند های خطی را میسازیم برای داده های یک و دو و آن هارا آموزش میدهیم.



تابع زیر نیز برای محاسبه دقت بر حسب داده و مدل میباشد

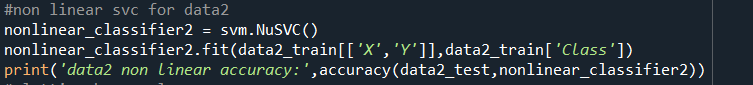


دقت دسته بند برای داده 1 و 2 را در حالت خطی مشاهده میکنیم



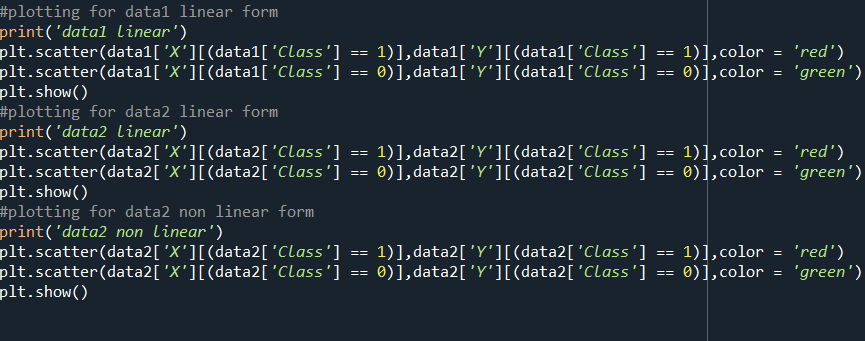


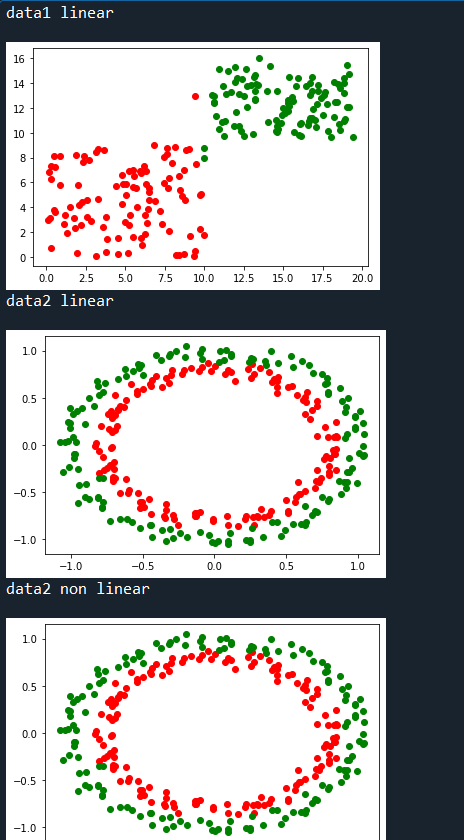
حال برای داده دوم یک دسته بند غیر خطی تعریف میکنیم و دقت آن را محاسبه میکنیم





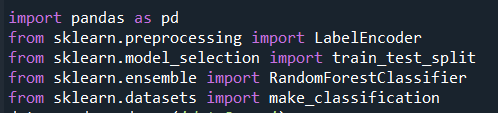
کد زیر نیز برای plot کردن داده هاست



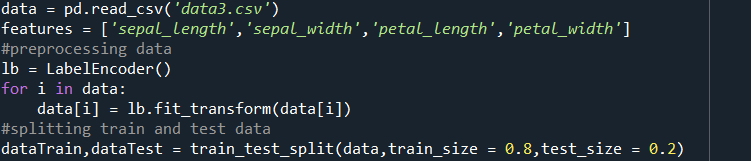


سوال 2)

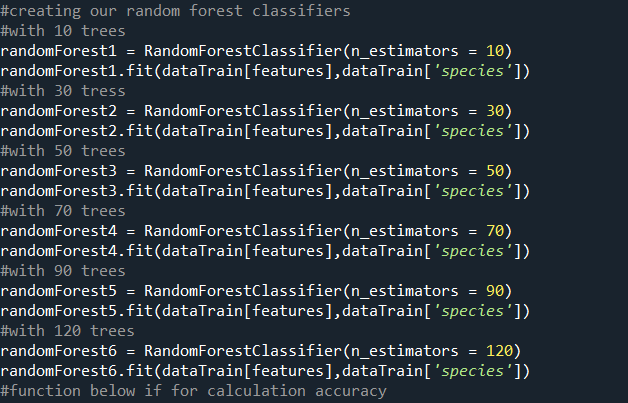
در ابتدا کتاب خانه های مورد نیاز را import میکنیم



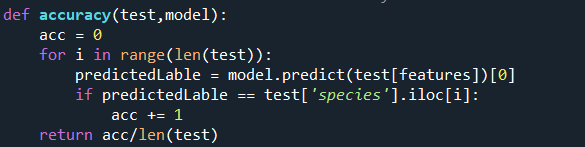
داده را میخوانیم و اسم ویژگی های آن را در آرایه نگهداری میکنیم و بعد از آن پیش پردازش داده را انجام میدهیم و داده های آموزش و تست را جدا میکنیم



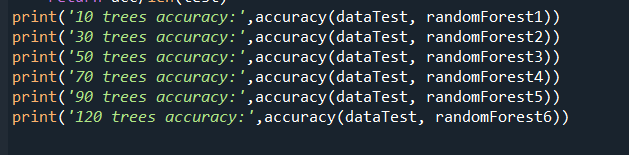
حال جنگل ها را با تعداد درخت متفاوت میسازیم

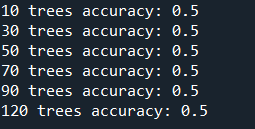


تابع زیر نیز برای محاسبه دقت میباشد



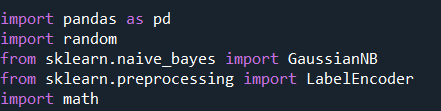
و دقت را برای تمامی جنگل ها محاسبه میکنیم و مشاهده میکنیم به ازای هر تعداد درخت دقت یکسانی میدهد.



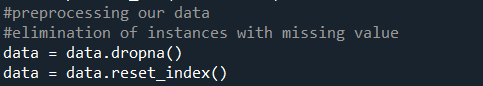


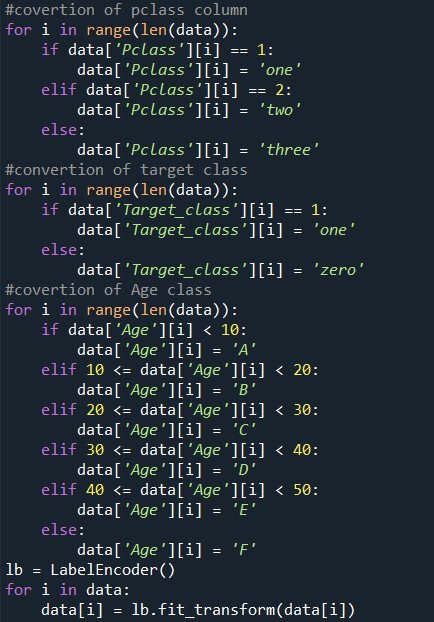
سوال 3)

اول کتاب خانه های مورد نظر را import میکنیم

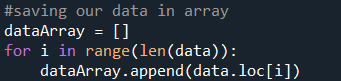


داده را میخوانیم و پیش پردازش های لازم را انجام میدهیم(برای مدیریت داده های missing سطر هایی که در آن missing value میباشد را حذف میکنیم.)

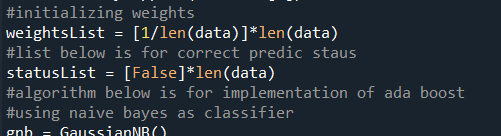




داده هارا به ترتیب در آرایه ذخیره میکنیم.



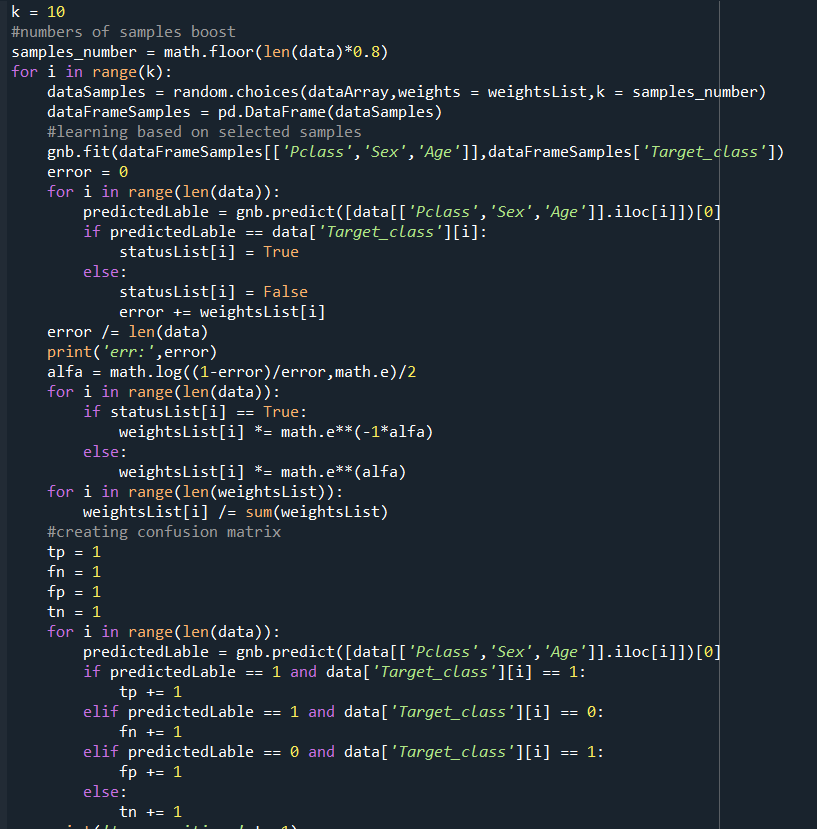
وزن های هر داده و وضعیت آنکه آیا درست پیش بینی شده اند یا خیر را در یک آرایه قرار میدهیم

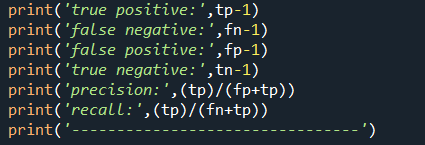


برای دسته بندی داده ها را در هر بخش نیز از بیض ساده استفاده میکنیم



الگوریتم زیر نیز پیاده سازی adaboost میباشد که در هر مرحله از آن confusion matrix را نمایش میدهیم





در شکل زیر نیز یکی از confusion matrix ها اورده شده است.

