



دانشگاه تهران

پردیس دانشکده های فنی

دانشکده مهندسی برق و کامپیوتر

یادگیری ماشین

تمرین دوم

امیرحسین عباسکوهی

۸۱۰۱۹۷۵۳۹

استاد ابولقاسمی

سوال ۶)

برای این بخش از تمرین قرار است از داده Iris استفاده کنیم و یک طبقه بند ساده را پیاده سازی کنیم. برای این منظور در ابتدا با استفاده از pandas داده ها را میخوانیم و متوجه میشیم که ۴ ویژگی بر اساس طول و عرض sepal و petal داریم.

حال ۶ نمودار بر اساس ۶ حالت از ترکیب ۴ ویژگی (انتخاب دو از ۴) رسم میکنیم.

بر اساس نمودار های به دست آمده باید تصمیم بگیریم کدام ویژگی برای طبقه بندی مناسب است. آن ویژگی مناسب است که کمترین در هم رفتگی داده را به اضای آن داشته باشیم. به همین دلیل من ویژگی Petal_width را انتخاب میکنم.

حال با نسبت هشتاد درصد و ۲۰ درصد تقسیم بندی به تمرین و تست را انجام میدهیم و سپس برای داده های تمرین نقطه میانگین را به کمک نامپای برای هر کلاس پیدا میکنیم. سپس برای هر نقطه فاصله از میانگین را محاسبه میکنیم و فاصله کمتر را معیار طبقه بندی قرار میدهیم.

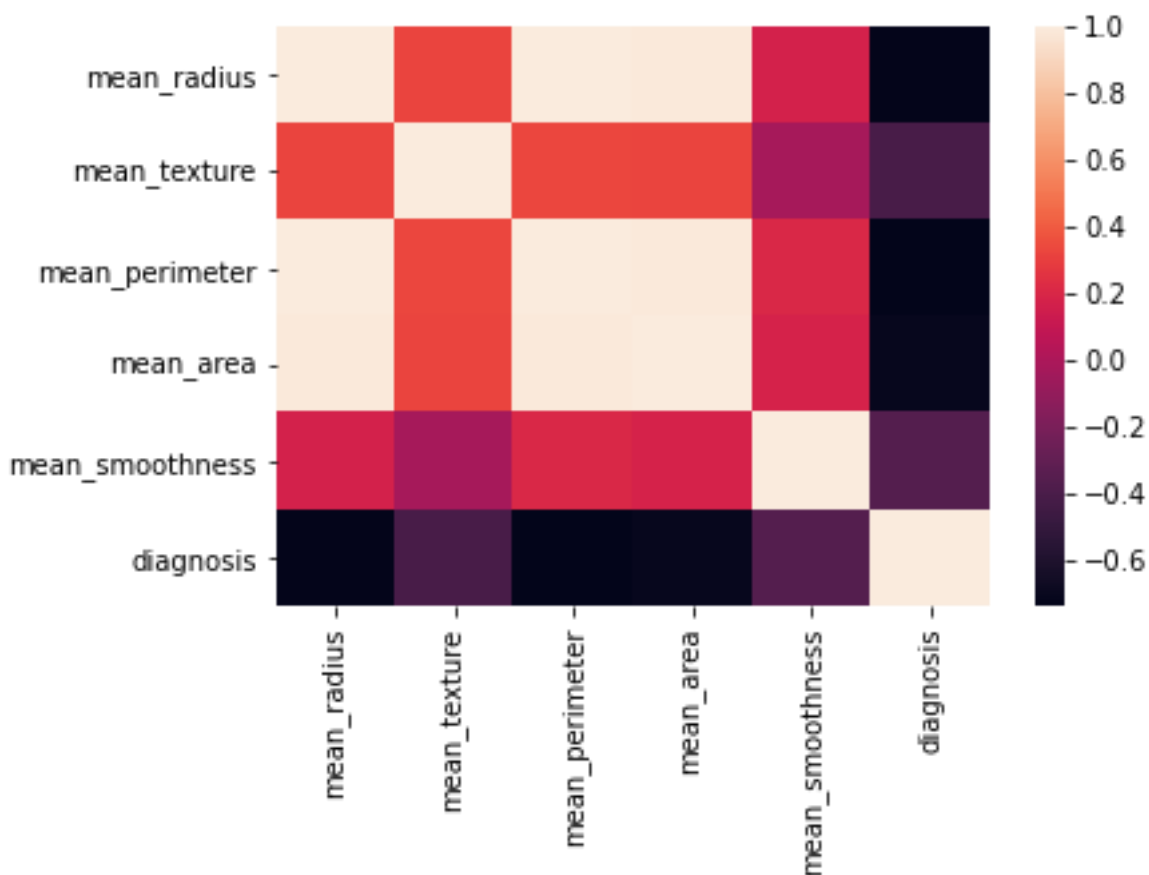
Confusion matrix و accuracy در نوت بوک ذکر شده است.

حال همین موارد را با استفاده از NearestCentroid از کتابخانه sklearn انجام میدهیم. نتایج این مورد هم در نوت بوک ذکر شده است.

در نهایت هم با استفاده از roc curve از sklearn و استفاده از تکنیک one vs all نمودار های roc را رسم میکنیم.

سوال ۷)

برای این نمودار در ابتدا به بررسی داده می پردازیمو در ابتدا نمودار correlation را رسم میکنیم. این نمودار را در اینجا میبینیم:



بر اساس نمودار کسینیم که mean_perimeter و mean_area هر دو در حال توصیف همان ویژگی mean_radius هستند. بر این اساس این دو ویژگی را از آنجا که اطلاعات خاصی به ما نمیدهند حذف میکنیم.

همینطور normalization داده ها را انجام میدهیم.

در مرحله بعد Bayesian classifier را پیاده سازی میکنیم که در واقع مقادیر ماتریس کوواریانس، احتمال اولیه و میانگین را ست میکنیم سپس برای دو کلاس g function را محاسبه میکنیم و بر اساس مقایسه مقادیر آن کلاس بندی میکنیم.

در مرحله بعد هم از GaussianNB در sklearn استفاده میکنیم.

طبقه بندی کننده optimal bayes , طبقه ای را انتخاب می کند که یک احتمال posterior بیشتری برای وقوع دارد (به طوری که برآورد درستی ما کسیمم نامیده می شود).

در Naïve Bayes اما با نگاه به توزیع تجربی و با فرض استقلال طبقه‌بندی را تخمین می‌زنند.

سوال ۸)

در این قسمت قرار است logistic regression را با استفاده از تابع داده شده انجام دهیم. برای این منظور در ابتدا ویژگی‌های x^2 ، y^2 و xy را اضافه می‌کنیم. در این صورت اگر یک linear regression با ویژگی‌های اخیر تمرین دهیم و در نهایت تابع logit را استفاده کنیم به logistic regression بر اساس تابع داده شده رسیده ایم.

تمرین داده به سادگی با استفاده از descending انجام شده است و پیچیدگی خاصی ندارد. در نهایت برای رسم نمودار هم از contour و تخمین طبقه بند استفاده می‌کنیم که تصویر مرز در نوت بوک موجود است.