

به نام خدا

پروژه : درخت تصمیم

محمد حسین میرزائی 99522158

کد های این پروژه برای پیاده سازی بهتر در colab پیاده شده است .

از کتاب خانه های `numpy` , `pandas` , `shutil` استفاده کرده ایم .

شامل 2 قسمت پیاده سازی برای مطالب جزوه و دیگری برای `dataset` دیابت می باشد .

از کتاب خانه `pandas` برای گرفتن دیتا استفاده شده است ، و با دستورات مربوط به `data frame` داده ها چاپ می شود .

`Class node` برای پیاده سازی گره استفاده می شود و دارای `constructor` برای گرفتن اطلاعات گره می باشد .

`Class DecisionTreeClassifier` بخش اصلی کد ما را شامل می شود که توابع اصلی ما نیز در آن تعریف شده است که در ادامه به توضیح آن می پردازیم .

در قسمت اصلی اطلاعات فایل را می گیریم و به 2 بخش `test` و `train` تقسیم و درخت تصمیم نظیر را می سازیم .

حال داده های `test` را برای `fit` می فرستیم و یک `dataset` کلی را برای `build tree` می فرستیم .

حال با توجه به dataset تعداد sample و feature ها را بدست می آوریم .

یک سری شروط توقف برای درخت در نظر می گیریم که اگر تعداد sample ها از مقداری کمتر بود و یا عمق درخت از مقداری بیشتر بود ، رشد درخت متوقف تا پیاده سازی راحت تر شود .

حال از این تابع به تابع split رفته و بهترین تقسیم بندی را انتخاب می کنیم و برای این کار روی اعضای ورودی تابع unique را صدا می زنیم و با استفاده از threshold بازه ای برای تقسیم بندی پیدا می کنیم و درخت را با توجه به threshold جدا سازی می کنیم و این جدا سازی بر اساس information gain می باشد ، که IG نیز از طریق entropy و Gini و فرمول بیان شده در جزوه را پیاده می شود .

برای قسمت برگ نیز برای پیاده سازی به قسمت calc\_leaf\_val رفته و مقدار polarity value را برای آن گره بدست می آوریم ، به روش انتخاب بیشترین تعداد برای آن گره در نظر می گیریم .

حال داده های train را در نظر می گیریم و در نهایت نیز با درخت ساخته شده جواب نهایی این ستون را predict می کنیم ، و با توجه به مقدار predict جواب نهایی را به تقریب به دست می آوریم .

Alt	Bar	Fri	Hun	Pat	Price	Rain	res	Type	Est	Goal	
0	1	0	0	1	1	2	0	1	0	1	
1	1	0	0	1	2	0	0	0	1	2	0
2	0	1	0	0	1	0	0	0	2	0	1
3	1	0	1	1	2	0	1	0	1	1	1
4	1	0	1	0	2	2	0	1	0	3	0
5	0	1	0	1	1	1	1	1	3	0	1
6	0	1	0	0	0	0	1	0	2	0	0
7	0	0	0	1	1	1	1	1	1	0	1
8	0	1	1	0	2	0	1	0	2	3	0
9	1	1	1	1	2	2	0	1	3	1	0
10	0	0	0	0	0	0	0	0	1	0	0
11	1	1	1	1	2	0	0	0	2	2	1



Pregnancies Glucose BloodPressure SkinThickness Insulin BMI DiabetesPedigreeFunction Age Outcome



0	0	1	1	0	0	1	0	1	1
1	0	1	1	0	0	1	0	1	0
2	0	1	1	0	0	0	0	1	1
3	0	1	1	0	0	1	0	0	0
4	0	1	0	0	0	1	1	1	1
5	0	1	1	0	0	1	0	0	0
6	0	1	1	0	0	1	0	0	1
7	1	1	0	0	0	1	0	0	0
8	0	1	1	0	1	1	0	1	1
9	0	1	1	0	0	0	0	1	1
10	0	1	1	0	0	1	0	0	0
11	1	1	1	0	0	1	0	1	1
12	1	1	1	0	0	1	1	1	0
13	0	1	1	0	1	1	0	1	1
14	0	1	1	0	0	1	0	1	1
15	0	1	0	0	0	1	0	1	1
16	0	1	1	1	0	1	0	1	1
17	0	1	1	0	0	1	0	1	1
18	0	1	0	0	0	1	0	1	0
19	0	1	1	0	0	1	0	1	1
20	0	1	1	0	0	1	0	0	0
21	0	1	1	0	0	1	0	1	0
22	0	1	1	0	0	1	0	1	1
23	1	1	1	0	0	1	0	0	1