

برای قسمت اول ابتدا accuracy هر فیچر را با الگوریتم طبقه بندی ID3 محاسبه میکنیم که در اینجا با الگوریتم K-fold برای 5 حالت accuracy را محاسبه میکنیم سپس میانگین این 5 مقدار را برای هر فیچر ست میکنیم.

در مرحله بعد ماتریس correlation را به دست می آوریم و برای همه مقادیر قدر مطلق را حساب میکنیم.

سپس میانگین correlation را برای هر فیچر حساب میکنیم :

در اینجا با توجه به نحوه پیاده سازی ، برای به دست آوردن f1، من میانگین را به جای max برای هر فیچر حساب کردم.

حالا دو متریک accuracy , mean_correlation برای هر فیچر داریم که من 17 فیچر در نظر گرفته ام.

برای به دست آوردن f1 ابتدا correlation را عکس میکنیم تا بیشترین مقدار بهترین باشد مانند accuracy و برای هر فیچر مقدار f1 را حساب میکنیم.

حالا این مقادیر را سورت میکنیم و بر اساس این مقدار و تعداد فیچری که میخواهیم، دیتا اصلی را برمیگردانیم.

در مرحله بعد با استفاده از الگوریتم kmeans داده ها را خوشه بندی میکنیم و هر خوشه را با استفاده از یک classifier آموزش می دهیم و سپس یک حلقه بر روی دیتا تست میزنیم و predict_label هر داده تست را به دست می آوریم و در نهایت دقت آن را محاسبه میکنیم که البته برای این دیتا به دلیل اینکه دیتا تمیزی هست به طور واضح نمیتوانیم بهتر شدن دقت را ببینیم و در چند حالت خاص با تغییر تعداد کلاستر ها و فیچر ها این اتفاق رخ می دهد.

در مرحله آخر هم برای یک حالت سه کلاسه (AB-CD-E) و سه حالت دو کلاسه

(AB-CD , AB-E, CD-E) دقت را اندازه گیری میکنیم.