

عرفان ابراهیمی سویز

۹۹۵۲۲۶۶

کد اجرایی برنامه با توضیحات:

۱- کد اول برای دیتاست رستوران است که تفاوتی که با کد دیگر دارد به این صورت است که اول دیتا را تغییر می دهد به این صورت که:

برای yes یک می گذارد و برای no صفر می گذارد

برای اجرا گرفتن کافی است run all را در جویپتر بزنید و box آخر درخت را به ما نمایش می دهد

برای کد دوم که daibet است به همین شکل اجرا می شود.

نمایش درخت:

۲- نمایش درخت به این صورت است که در هر عمق یک tab جلو می رود

```
threshold:[0, 1, 2] and feature:Pat and gain: 0.3748900964125389
  value: 0
  value: 1
    threshold:[1, 2, 3] and feature:Type and gain: 0.17441604792151594
      value: 0
        threshold:[1, 2] and feature:Est and gain: 0.6931471805599453
          value: 1
          value: 0
        threshold:[0, 1] and feature:Hun and gain: 0.6931471805599453
          value: 0
          value: 1
```

Threshold مقدار هایی دارد که با آن بازه ها را تقسیم کردیم برای مثال در راس درخت کمتر از صفر و بین یک و صفر و بین یک و دو و بیشتر از دو تقسیم کردیم و از آنجایی که بیشتر از دو نداشتیم و کمتر از صفر نداریم کلا به سه دسته تقسیم شده است پس کلا سه تا بچه داریم که feature آن برابر با pat است.

-۳

مسیر انجام کار:

اول باید دیتا را بخوانیم و آن ها را به عدد تبدیل کنیم همان طور که در مثال اسلاید ها این کار را کردیم

حال باید یکی یکی feature ها را انتخاب کنیم و آن ها را به تعداد دسته هایی که لازم داریم تقسیم کنیم. تقسیم کردن به این صورت است که value ها هر feature را منحصر به فرد می کنیم و آن را sort می کنیم و در نهایت به تعداد بچه هایی که لازم داریم تقسیم می کنیم مثلاً به سه دسته تقسیم می کنیم حال entropy هر کدام را حساب می کنیم و بعد information gain را حساب می کنیم و این کار را برای تمام feature ها انجام می دهیم و بهترین را انتخاب می کنیم و به همین صورت ادامه می دهیم تا درخت درست شود .

حال باید درخت را امتحان کنیم به این صورت که یک مقدار از دیتا را برای امتحان کردن استفاده می کنیم در این صورت می توانیم درخت را امتحان کنیم.

بعد باید درخت را بکشیم به همان صورتی که گفته شد درخت رسم می شود.

اولین چالش بازگشتی بودن تابع بود که باعث می شد تا کد زدن آن سخت شود.

بزرگ ترین چالش آن تقسیم بندی دیتای یک feature بود زیرا بهترین حالت برای بدست آوردن دیتا به این صورت است که باید information gain یک feature را با threshold های مختلف حساب کنیم در این صورت می توانیم بهترین مقدار را بدست بیاوریم اما به این صورتی که ما می خواستیم به چند تا بچه تقسیم کنیم تعداد حالت ها زیاد می شد در این صورت ما نمی توانستیم بهتر حالت ممکن را بدست بیاوریم به همین علت دقت کار پایین می آمد.

و همین طور قسمی کردن دیتا ها اگر به درستی صورت نمی گرفت باعث می شد که gain کمتر از صفر شود و یا درخت به درست چاپ نشود.

برای دیتاست diabetes چهار بچه را در نظر می گیریم و دقت آن برابر با ۷۴ درصد می شود که بخشی از درخت به این صورت است و عمق نهایتا ۱۰۰ تا است

```
threshold:[96.0, 129.0, 162.0] and feature:Glucose
threshold:[3.0, 7.0, 10.0] and feature:Pregnancies
threshold:[24.7, 28.9, 35.0] and feature:BMI
value: 0
value: 0
value: 0
threshold:[48.0, 72.0, 105.0] and feature:Insulin
threshold:[35.8, 37.4, 38.2] and feature:BMI
value: 0
value: 0
value: 1
threshold:[0.194, 0.503] and feature:DiabetesPedigreeFunction
value: 0
value: 0
threshold:[22.0, 23.0, 43.0] and feature:Age
value: 0
value: 0
value: 1
value: 1
threshold:[26.0, 31.0, 37.0] and feature:Age
value: 0
threshold:[23.0, 54.0, 70.0] and feature:Insulin
value: 0
value: 0
value: 0
value: 1
threshold:[62.0, 74.0, 80.0] and feature:BloodPressure
value: 0
threshold:[62.0, 64.0] and feature:BloodPressure
value: 0
value: 1
value: 0
threshold:[3.0, 5.0, 6.0] and feature:Pregnancies
value: 0
value: 1
value: 0
value: 0
threshold:[27.6, 30.1, 33.3] and feature:BMI
value: 0
value: 1
value: 0
threshold:[75.0, 88.0, 90.0] and feature:Glucose
value: 0
value: 0
```

حال بچه ها را به دو تا کاهش می دهیم و دقت به 70 درصد کاهش پیدا می کند.

```

threshold:[129.0] and feature:Glucose
  threshold:[7.0] and feature:Pregnancies
    threshold:[31.6] and feature:BMI
      threshold:[37.0] and feature:Age
        threshold:[3.0] and feature:Pregnancies
          value: 0
            threshold:[26.0] and feature:BMI
              threshold:[103.0] and feature:Glucose
                value: 0
                  threshold:[27.0] and feature:Age
                    threshold:[48.0] and feature:Insulin
                      value: 0
                      value: 1
                    value: 0
                  threshold:[5.0] and feature:Pregnancies
                    threshold:[85.0] and feature:Insulin
                      value: 0
                      threshold:[110.0] and feature:Glucose
                        threshold:[32.0] and feature:SkinThickness
                          value: 0
                          value: 0
                        threshold:[0.368] and feature:DiabetesPedigreeFunction
                          value: 0
                          threshold:[6.0] and feature:Pregnancies
                            value: 1
                            threshold:[32.0] and feature:Age
                              threshold:[33.0] and feature:Age
                                value: 0
                                value: 1
                              value: 1
                            threshold:[27.6] and feature:BMI
                              threshold:[3.0] and feature:Pregnancies
                                threshold:[75.0] and feature:BloodPressure
                                  value: 0
                                  threshold:[80.0] and feature:BloodPressure
                                    value: 1
                                  value: 0
                                threshold:[114.0] and feature:Glucose
                                  threshold:[41.0] and feature:Age
                                    value: 0
                                    threshold:[0.722] and feature:DiabetesPedigreeFunction
                                      value: 1
                                    threshold:[76.0] and feature:BloodPressure
                                      value: 1
                                      threshold:[22.0] and feature:Insulin
                                        threshold:[115.0] and feature:Glucose
                                          value: 0
                                          value: 1
                                        value: 0

```

تعداد بچه ها را به ۸ تا افزایش داده شد و نهایت عمق به ۲۰۰ تا اما دقت به 67 کاهش پیدا کرد.

```

Output exceeds the size limit. Open the full output data in a text editor
threshold:[79.0, 96.0, 112.0, 129.0, 146.0, 162.0, 179.0] and feature:Glucose
  threshold:[23.0, 26.0, 33.0, 36.0, 38.0, 41.0, 46.0] and feature:Age
    value: 0
    value: 0
    threshold:[26.8, 31.0] and feature:BMI
      value: 0
    value: 0
    threshold:[0.0, 32.0] and feature:SkinThickness
      value: 0
      value: 1
    value: 0
    value: 0
    value: 0
    threshold:[52.0, 58.0, 65.0, 70.0, 75.0, 82.0, 88.0] and feature:BloodPressure
      value: 0
      value: 0
      threshold:[22.0, 23.0, 24.0, 28.0, 31.0, 33.0, 44.0] and feature:Age
        value: 0
        value: 0
        threshold:[21.2, 29.9, 38.0, 43.5] and feature:BMI
          value: 0
          value: 0
          value: 1
          value: 0
        value: 0
      ...
        value: 0
      value: 1
      value: 1
      value: 1

```

با تعداد بچه های ۴ تا و عمق ۷۰ تا به ۷۲ درصد افزایش پیدا کرد.

```

threshold:[96.0, 129.0, 162.0] and feature:Glucose
  threshold:[3.0, 7.0, 10.0] and feature:Pregnancies
    threshold:[24.7, 28.9, 35.0] and feature:BMI
      value: 0
      value: 0
      value: 0
      threshold:[48.0, 72.0, 105.0] and feature:Insulin
        threshold:[25.0, 29.0, 32.0] and feature:SkinThickness
          value: 0
          value: 1
          value: 0
          threshold:[25.0, 27.0] and feature:Age
            value: 0
            value: 1
          value: 0
          value: 0
          threshold:[0.347, 0.366, 0.674] and feature:DiabetesPedigreeFunction
            value: 1
            value: 0
            value: 0
          threshold:[26.0, 31.0, 37.0] and feature:Age
            value: 0
            threshold:[23.0, 54.0, 70.0] and feature:Insulin
              value: 0
              value: 0
            ...
              threshold:[0.222, 0.427, 0.702] and feature:DiabetesPedigreeFunction
                value: 1
                value: 0
                value: 1

```