

به نام خدا

گزارش کار و راهنمای اجرای

تمرین پیاده سازی دوم

احمد زعفرانی

97105985

درس: هوش مصنوعی

استاد عبدی

## راهنمای اجرا

پوشه **part1** مربوط به قسمت اول تمرین، و پوشه **part2** مربوط به قسمت دوم است. در هر دو پوشه، دو فایل پایتون قرار دارند؛ که برای اجرای برنامه باید فایل `part1.py/part2.py` را اجرا کنید.

در قسمت اول، برنامه ورودی ها را از کنسول دریافت می کند. ابتدا تعداد داده های آموزشی، سپس هر کدام از آنها را در خطوط جداگانه به برنامه بدهید. در هر نمونه، ویژگی ها باید به ترتیبی که در صفحه 26 اسلاید های درس آمده، در یک خط نوشته شوند (ویژگی ها با یک فاصله از هم جدا شوند). برنامه به املای ورودی ها حساس است، حرف اول کلماتی مانند **Yes** باید با حروف بزرگ نوشته شود (دقیقا مانند صفحه 26). سپس نمایش گرافیکی درخت نشان داده می شود. حال تعداد نمونه های آزمایشی توسط برنامه خواسته می شود و در نهایت به ازای هر کدام از این نمونه ها (نمونه ها با فرمت داده های آموزشی، در خطوط جداگانه به برنامه داده شوند)، اینکه درخت تصمیم نتیجه را به درستی پیش بینی کرده یا خیر، چاپ می شود. (دقت کنید ویژگی **goal** حتی هنگام وارد کردن نمونه های آزمایشی باید به برنامه داده شود). برای سهولت کار یک فایل `test.txt` در پوشه اول قرار داده شده است که در آن هر دوازده داده آموزشی در صفحه 26 جزوه به فرمت مورد انتظار برنامه نوشته شده است؛ می توانید آنها (یا تعدادی از آنها) را کپی کرده و درون کنسول **paste** کنید.

در قسمت دوم، ورودی ها از روی فایل `diabetes.csv` خوانده می شوند (دقت کنید که این فایل باید با فایل های اجرایی برنامه درون یک پوشه باشند). برای تعیین آنکه چه تعداد از آنها داده آموزشی باشند و چه تعداد داده آزمایشی، یک عدد تصادفی بین 50 و 90 مانند **x** در نظر می گیریم؛ و **x** درصد اول داده های فایل را به عنوان داده های آموزشی و بقیه را داده های آزمایشی فرض می کنیم. سپس درخت تصمیم متناظر با ورودی ها ترسیم می شود، و داده های آزمایشی در آن تست می شوند. در نهایت، درصد صحت پیش بینی های درخت تصمیم ایجاد شده، در خروجی نمایش داده می شود.

## جزئیات پیاده سازی

هر دو قسمت تمرین از توابع و کلاس های بسیار مشابهی استفاده می کنند؛ در واقع به جز نحوه ورودی و خروجی دادن و کلاس **Sample** که باید به اقتضای نوع و ویژگی داده ها متفاوت طراحی شود، تفاوت دیگری میان دو قسمت مختلف تمرین وجود ندارد. پیاده سازی الگوریتم ساخت درخت تصمیم، دقیقا مشابه با صفحه 30 جزوه صورت گرفته است.

درخت تصمیم در قسمت اول تمرین، به پیوست این گزارش آمده است. سطوح ابتدایی این درخت مشابه شکل سمت راست در صفحه 28 جزوه می باشد:

در قسمت دوم، هر کدام از داده های پیوسته را به 10 بازه تقسیم می کنیم. در این حالت، میانگین صحت پیش بینی های درخت، به 65 درصد می رسد.

درباره فرمت خروجی: متأسفانه به علت محدود بودن ددلاین تمرین، راه حل مناسبی برای نمایش درخت بصورت گرافیکی پیدا نکردم (کتابخانه هایی برای این کار در زبان پایتون وجود دارند، ولی زمان برای یادگیری کار با آنها کوتاه بود...); حتما این مشکل در زمان تحویل حضوری رفع خواهد شد.

اما هر گره درخت با فرمت زیر نمایش داده می شود (از پیمایش پیش ترتیب برای پیمایش درخت استفاده کرده ایم):

Value – branch – entropy – remainder – gain

برای مثال:

remainderEntropy = 0.000      entropy = 0.000      inBranch = insulin = >830      value = 1  
informationGain = 0.000

یعنی یک برگ از درخت تصمیم، خروجی 1 است (فرد به دیابت مبتلا است)، پدر این برگ ویژگی insulin بوده مقدار آن در بازه اعداد بزرگتر از 830 قرار داشته است. بی نظمی، بی نظمی باقیمانده و دستاورد اطلاعات نیز نشان داده شده اند.

یا مثالی برای قسمت اول:

remainderEntropy      entropy = 0.918      inBranch = patrons = Full      value = hungry  
informationGain = 0.252      = 0.667

یعنی یکی از گره های میانی درخت است، ویژگی hungry را باز نمایی می کند، پدر آن patrons بوده که هم اکنون در شاخه Full آن وارد شده ایم.

پایان