

پروژه درخت تصمیم
درس هوش مصنوعی
استاد عبدی

علی عطاریان – ۹۹۵۲۱۴۵۱

بخش اول) با ران گرفتن از کد `WaitingTrain.py` درخت مربوط به بخش اول که همان مثال جزوه است چاپ خواهد شد و در فایل `waiting.txt` ذخیره خواهد شد. (یک نمونه قبلا در `waiting1.txt` ذخیره شده است). این درخت از ویژگیهای `pat`، `est`، `bar` و `res` استفاده می کند. فرمت چاپ درخت به صورتی است که در هر نود ویژگی (`feature`) مورد بررسی، آنتروپی مربوط به آن نود و دست آورد حاصل از مشاهده ویژگی مذکور نمایش داده شده است. همچنین روابط فرزند و والد به صورت نقطه چین مشخص است.

بخش دوم) با ران گرفتن از کد `DiabetesTrain.py` درخت مربوط به بخش دوم بر اساس فایل دیابت چاپ خواهد شد و در فایل `diabetes.txt` ذخیره خواهد شد. (یک نمونه قبلا در `diabetes1.txt` ذخیره شده است). فرمت چاپ درخت همانند بخش قبل می باشد. دقت درخت با ۱۰ درصد داده آزمایشی، حدود ۶۷ درصد به دست آمد. برای ۲۰، ۳۰، ۴۰، ۵۰ درصد داده آموزشی متناظرا ۶۳، ۶۴، ۶۳، ۶۴ درصد دقت بدست خواهد آمد.

توضیحات) ابتدا کلیت درخت را با کمک [این ویدیوی آموزشی](#) پیاده سازی می کنیم. اما سه مشکل داریم. اول، نمایش درخت چگونه باشد؟ دوم، درخت دودویی است پس باید آن را به جای اینکه دارای فرزند چپ و راست باشد طوری تنظیم کرد که فرزندان (`children`) داشته باشد. سوم، نحوه تصمیم گیری برای انتخاب ویژگی بر اساس `threshold` است یعنی یکی از مقادیر آن ویژگی به عنوان آستانه انتخاب می شود و همه مقادیر در مقایسه با آن به فرزند چپ یا راست می روند اما ما می خواهیم بازه مقادیر ویژگی را تقسیم بندی کنیم. مشکل اول را با استفاده از کتابخانه `treelib` حل می کنیم یعنی همزمان با درخت تصمیم خودمان، درخت مشابه را بر اساس کتابخانه مذکور می سازیم تا بتوانیم از امکان `tree.show()` استفاده کنیم. مشکل دوم با ایجاد یک لیست فرزندان و قرار دادن `loop` در جاهای لازم حل می شود. برای حل مشکل سوم یک `property` به نام `splitter` به کلاس درخت تصمیم اضافه می کنیم تا تعیین کنیم هر بازه `max-min` در هر ویژگی به چند قسمت، تقسیم شود. در این مسیر به ارورهای زیادی از جمله `list assignment, the requested array has an inhomogeneous shape`

the truth value of an array is ambiguous ,index out of range و... برمی خوریم

که هر کدام را به نحوی پشت سر می گذاریم (:)

لازم به ذکر است که درخت تصمیم ساخته شده را یکبار با دیتاست از پیش آماده سرطان سینه نیز آزمایش می کنیم که شکل درخت آن در `train.txt` موجود است و دقت آن ۶۴ درصد است. البته همچنان دقت درخت بسیار پایین است و مشکلاتی در شکل درخت هم وجود دارد که قابل بررسی و بهبود است.