

به نام خدا

تمرین عملی ۱

به موارد زیر توجه کنید:

- برای این تمرین نمیتوانید از کتابخانه های آماده درخت تصمیم استفاده کنید.
- این تمرین را میتوانید در گروه های دو نفره انجام دهید.
- این تمرین تحويل حضوری دارد پس هر دو اعضای گروه باید بر کل تمرین مسلط باشند.
- تحلیل درست و کامل با معیار های Gini و Information gain نمره امتیازی دارد.

توضیحات دیتاست: این دیتاست شامل ۸ ویژگی و شما باید مشخص کنید که آیا این مدرسه را پیشنهاد میدهید یا خیر؟

این ۸ ویژگی در ۳ زمینه تقسیم بنده شده اند.

۱. وضعیت کاری والدین و ویژگی های مدرسه

• Parent's occupation

• Child's nursery

۲. ساختار خانواده و وضعیت مالی

• Form of the family

• Number of children

• Housing conditions

• Financial standing of the family

۳. وضعیت اجتماعی و سلامت خانواده

• Social conditions

• Health conditions

۱. با الگوریتم ID3 - و مجموعه دادگان آموزش (۸۰ درصد داده ها) داده شده یک درخت تصمیم بسازید. همانطور که در درس اشاره شد، برای انتخاب بهترین ویژگی در هر مرحله باید از information gain استفاده کنید. میزان صحت طبقه بندی را روی مجموعه train محاسبه کنید.)
راهنمایی: در صورتی که پس از استفاده از تمام ویژگیها، کلاس مشخص نمیشود، همانطور که در درس اشاره شد باید از رای گیری بین داده ها برای برچسب زنی انتهای درخت استفاده کنید.)

- یکبار ۵۰ و بار دیگر ۷۵ درصد دادگان آموزش را بصورت تصادفی انتخاب کنید و درخت را با آن آموزش دهید (یکبار با عمق ۶ و بار دیگر با عمق ۸). سپس روی کل دادگان تست آن را تست کنید و در نهایت صحت طبقه بندی را روی دادگان آموزش و تست به همراه سائز درخت گزارش کنید (یک بار با انتروپی و بار دیگر با ضریب جینی^۱). چه تفاوتی مشاهده میکنید؟

- **امتیازی:** ۷۰ درصد دادگان را بصورت تصادفی به عنوان دادگان آموزش انتخاب کنید و ۲۰ درصد را به عنوان دادگان validation و ۱۰ درصد را به عنوان دادگان test. سپس به کمک دادگان آموزش با ID3 درخت تصمیم را بسازید و با دادگان validation عمل هرس را انجام دهید. (تا دو عمق). منحنی تغییر خطای طبقه بندی را روی هر ۳ مجموعه train ، validation و test را به ازاء تعداد گره های مختلف درخت بکشید.

- عمل هرس کردن را به صورت تصادفی نیز انجام دهید، اگر در حالتی بهتر از مورد بالا به جواب رسیدید گزارش کنید.

¹ Gini index

موعد تحويل : ٦ آذر ١٤٠١
روش تحويل : آپلود در courses

موفق باشید

سیدین