

به نام خدا



دانشکده علوم کامپیوتر

هوش مصنوعی و سیستم های خبره

پروژه اول (درخت تصمیم)

دکتر آرش عبدی

زمستان ۱۴۰۲

طراحان:

طاها موسوی

سید ماکان حاجی سید جوادی

پروژه اول درس هوش مصنوعی و سیستم های خبره

- در صورت وجود هرگونه ابهام در پروژه به طراحان پیام دهید.
- علاوه بر آپلود داخل کوئرا، پروژه ها به صورت شفاهی ارائه گرفته میشوند؛ بنابراین تسلط شما هنگام ارائه شفاهی نسبت به گزارش پروژه اهمیت بیشتری دارد. برای توضیح بیشتر، اگر پروژه ۱۰۰ نمره داشته باشد، نمره ارائه شفاهی ۹۰ و نمره گزارش، ۱۰ خواهد بود؛ بنابراین سعی کنید قبل از ارائه شفاهی نسبت به کد خود و مبحث پروژه آمادگی لازم را داشته باشید.
- با توجه به تنظیم شدن ددلاین توسط خود شما امکان تمدید وجود ندارد.
- زبان برنامه نویسی پیشنهادی پایتون است ولی مجاز به استفاده از هر زبان دلخواهی هستید. نمره هر زبان با هم برابر است.
- پروژه تک نفره است لذا از هرگونه تقلب در انجام پروژه خودداری کنید.

آیدی تلگرام طراحان:

- @TahaMusvi
- @ab2khiar

هدف پروژه: هدف از این پروژه استفاده از درخت تصمیم برای طبقه بندی مشتریان بانک بر اساس ویژگی های آنها و پیش بینی اینکه آیا آنها سپرده بلند مدت می سازند یا خیر. این پروژه از این جهت مفید است که میتواند به بانک کمک کند تا مشتریان سودده را شناسایی کند و با کمپین های بازاریابی آنها را جذب کند.

شرح پروژه:

۱. تجزیه و تحلیل مجموعه داده:

با تجزیه و تحلیل مجموعه داده ارائه شده که مربوط به پیش بینی تهیه سپرده بلندمدت کاربران بانک است، شروع کنید. ابتدا در صورت وجود داده های ناقص، آنها را بهبود دهید یا حذف کنید (داده های ناقص داده هایی هستند که یک یا چند ویژگی آنها در دسترس نباشد)؛ سپس ویژگی های موجود را طبقه بندی کنید.

۲. پیاده سازی الگوریتم درخت تصمیم

کلاس های مورد نیاز این پروژه، حتما توسط خودتان پیاده سازی شوند و نه کتابخانه های آماده!

در الگوریتم خود از شاخصه های آنتروپی و Gini-index استفاده کنید.

۳. آموزش درخت تصمیم

درخت تصمیم خود را با کمک مجموعه داده ای که در اختیار شما قرار داده شده است آموزش دهید. توجه کنید که این مجموعه داده هم برای آموزش، هم ارزیابی تنظیم شده

است. بنابراین باید براساس مطالب درسی گفته شده و آزمون و خطا، میان داده های آموزشی و ارزیابی، تعادل مناسب را برقرار کنید.

۴. ارزیابی مدل

مدل خود را با استفاده از تکنیک های cross-validation و محاسبه خطا (مانند MSE) مورد بررسی قرار دهید و سعی کنید به حداقل دقت مطلوب (حدوداً ۷۰ درصد) برسید. هرچند عملکرد کلی شما و درستی الگوریتم از اهمیت بیشتری برخوردار است.

۵. نمایش درخت تصمیم

درخت تصمیم خود را با کمک کتابخانه های موجود می توانید نمایش دهید. دقت کنید که خروجی باید واضح باشد و بیانگر درستی درخت شما باشد.

۶. تحلیل و بررسی نتایج

درخت تصمیم ساخته شده بر اساس دو شاخصه آنتروپی و Gini-index را باهم مقایسه کنید. با استفاده از تکنیک های موجود در اسلایدها یا اینترنت می توانید درخت خود را بهبود ببخشید.

میزان دقت درخت شما باید در هنگام اجرا نمایش داده شود همچنین نتایج به دست آمده در فایل توضیحات تحلیل شود.

اگر از ایده ی جدید و خلاقانه ای استفاده کرده اید در فایل توضیحات به آن اشاره کنید. مثلاً ممکن است این خلاقیت باعث افزایش دقت درخت شما شود.

خروجی مورد انتظار:

- کد نهایی برنامه به همراه توضیحات لازم برای اجرای کد
- گزارشی مختصر از مراحل انجام کار و چالش هایی که مواجه شدید، اجراهای گرفته شده و روند پیشرفت پروژه، توضیحاتی در مورد تفاوت دو شاخص آنتروپی و Gini-index و

توضیحاتی در مورد معیار ارزیابی دقت نیز باید ارائه شود. (برای مثال آیا بیش برآزش¹ داشته اید؟ ایده ای برای افزایش دقت دارید؟ حتی اگر پیاده سازی نکردید آنرا ذکر کنید).

- هر گونه تحلیل اضافه مفید و خلاقیت می تواند نمره امتیازی داشته باشد.

موفق باشید!

¹ overfit