

به نام خدا



هوش مصنوعی و سیستم های خبره

پروژه اول - درخت تصمیم

دکتر آرش عبدی

پاییز 1403

طراحان :

محمدصادق نعمت پور

نیایش خانی

- در صورت وجود هر گونه ابهام در سوالات تنها به طراح آن سوال پیام دهید.
- با توجه به تنظیم شدن ددلاین تمارین توسط خود شما امکان تمدید وجود ندارد.
- زبان برنامه نویسی و قالب تمپلیت پایتون است ولی برای تمرین های اول می توانید از **C#** نیز استفاده کنید.
- کل محتوای ارسالی را زیپ کرده و نام آن را شماره دانشجویی خود قرار دهید.
- نیازی به نوشتن داک نیست.
- به سوالات انتهایی با دقت و به صورت کامل پاسخ دهید.
- انجام تمرین تک نفره است. لطفا به تنهایی انجام شود، در غیر اینصورت نمره منفی در نظر گرفته خواهد شد.

آیدی تلگرام طراحان :

[@msnp1381](https://t.me/msnp1381)

[@mainlynia](https://t.me/mainlynia)

پروژه اول هوش مصنوعی و سیستم های خبره: درخت تصمیم

هدف پروژه: پیاده سازی درخت تصمیم با استفاده از آنتروپی و Gini index برای دیتاست های ذکر شده

شرح پروژه:

1. تجزیه و تحلیل مجموعه داده:

با تجزیه و تحلیل مجموعه داده های ارائه شده شروع کنید. Feature های موجود را گسسته سازی کنید، مقادیری که موجود نیستند را با مقداری مناسب جایگزین کنید (و یا می توانید آن سطر را حذف کنید)، مواردی که به صورت عدد نیستند را به عدد تبدیل کنید و برای گسسته سازی ورودی های از نوع پیوسته یا ورودی های دارای مقادیر خیلی زیاد بازه های عددی در نظر بگیرید. یک ایده این است که بازه مینیمم تا ماکزیمم اعداد در مجموعه آموزشی را به تعدادی بازه مساوی تقسیم کنید و دو بازه اضافی هم برای مقادیر کمتر از مینیمم و بیشتر از ماکزیمم در نظر بگیرید. همچنین میتوانید ایده های دیگری را نیز برای گسسته سازی ورودی های پیوسته ارائه دهید و آنها را امتحان کنید.

2. پیاده سازی الگوریتم درخت تصمیم:

کلاسها و الگوریتم درخت تصمیم مورد نظر خود را پیاده کنید. در نوت بوک هایی که در اختیار شما قرار گرفته اند، توابع آماده پیاده سازی هستند و کافی است با خواندن توضیحات موجود در کد، آنها را برای هر کدام از دیتاست ها تکمیل کنید. (main-crona.ipynb و main-titanic.ipynb) اگر در پیاده سازی این الگوریتم خلاقیت و نوآوری داشته باشید، نمره امتیازی خواهد داشت؛ مثلاً الگوریتم شما به صورت تطبیقی (adaptive) باشد و

یک بار الگوریتم خود را با آنتروپی و بار دیگر با استفاده از Gini index پیاده سازی کنید.

3. آموزش درخت تصمیم (training):

درخت تصمیم خود را با استفاده از دو دیتاست های زیر train کنید. می توانید از 2000 داده اول برای این کار استفاده کنید. (توجه داشته باشید که الگوریتم شما باید برای هر دو دیتاست به صورت جداگانه اجرا شود.)

Dataset#1: [titanic](#) | Dataset#2: [corona test](#)

4. ارزیابی مدل:

مدل درخت تصمیم خود را با استفاده از تکنیک های cross-validation ارزیابی کنید و نتایج را نشان دهید. از میان داده هایی که برای train استفاده نکرده اید به صورت تصادفی برای test انتخاب کنید.

5. تحلیل و بررسی نتایج:

درخت تصمیم بر اساس آنتروپی و Gini index را با یکدیگر مقایسه کنید. سعی کنید با استفاده از تکنیک های مختلف درخت تصمیم خود را بهبود بخشید و نتایج به دست آمده را تحلیل کنید. میزان دقت درخت خود را نمایش دهید.

اگر از ایده جدیدی در هر قسمت پروژه استفاده کردید (مثلا استفاده از روش جدیدی برای گسسته سازی داده ها)، نشان دهید که این کار چه تغییری در درخت تصمیم شما ایجاد کرده است و آیا باعث بهبود آن شده است یا خیر.

✓ خلاقیت شما برای افزایش دقت درخت مثل افزایش داده های آموزشی یا هر گونه انتخاب هوشمندانه از میان آنها، روش های جدیدتر و حرفه ای تر گسسته سازی و یا حتی فعالیت های اضافه تر حرفه ای مانند تحلیل های آماری جداگانه از فیچرها، Data cleaning یا Feature engineering و .. می تواند نمره امتیازی داشته باشد.

✓ در نظر داشته باشید برای پیاده سازی درخت تصمیم نباید از توابع آماده استفاده کنید. لذا فرمول آنتروپی و Gini index تابع خود درخت تصمیم (همانند توابع بازگشتی و فرآیند درخت سازی و ...) را باید خودتان پیاده کنید.

✓ استفاده از توابع آماده تنها برای بخش های دیگر مانند خواندن اکسل، احیانا نمایش گرافیکی خروجی درخت (در صورت علاقه)، نمایش دقت خروجی و ... بلامانع است.

ساختار پروژه و نحوه تحویل

```
└─ corona
|   └─ __init__.py
|       └─ main-corona.ipynb
└─ titanic
|   └─ __init__.py
|       └─ main-titanic.ipynb
└─ DecisionTree_4031.docx
└─ decision_tree.py
└─ questions.docx
└─ README.md
└─ requirements.txt
```

وظیفه شما تکمیل decision_tree.py که نشان دهنده ساختار درخت تصمیم است.

ساختار دو نوتبوک main-titanic.ipynb و main-corona.ipynb کاملاً شبیه است اما در مصورسازی و پیش پردازش تفاوت‌هایی وجود دارد. که در این‌ها دیتاست‌های مرتبط لود و انجام می‌شود.

برای بارگذاری پروژه حتماً REAME.md را مطالعه کنید.

آنچه تحویل داده می‌شود:

1. کد اجرایی برنامه با توضیحات لازم برای اجرا
2. پاسخ به سوالات قرار گرفته در داک برای هر دیتاست (questions.docx)
3. هر گونه تحلیل اضافه مفید و خلاقیت 😊 (می‌تواند نمره امتیازی داشته باشد).