



دانشگاه اصفهان
دانشکده مهندسی کامپیوتر
گروه مهندسی هوش مصنوعی

هوش مصنوعی و سیستم های خبره
مجموعه دوم تمرین

استاد: دکتر کارشناس

پاییز ۱۴۰۱

فهرست

۳	سوال شماره ۱
۷	سوال شماره ۲
۹	سوال شماره ۳
۱۳	سوال شماره ۴

سول شماره ۱

استخدام



الف) معرفی کلی مسئله

مراحل استخدام در شرکتی به این نحو است که در زمان معین از متقاضی خواسته میشود آزمونی شامل ده مجموعه سوال را انجام دهد و پس از اتمام این آزمون تصحیح شده و نمره‌ای بین ۰ تا ۲۰ داده میشود (موضوعات این مجموعه سوالات عبارتند از ۱. ساختمان داده ۲. طراحی الگوریتم ۳. یادگیری ماشین ۴. یادگیری عمیق ۵. پایگاه داده ۶. آشنایی با سخت افزار کامپیوتر ۷. آشنایی با ساختار سیستم عامل ۸. شبکه های کامپیوتری ۹. پردازش سیگنال دیجیتال ۱۰. آمار و جبر خطی است)

پس از پایان آزمون پروژه‌ای مرتبط به حوزه استخدامی داده شده و پس از بررسی پروژه نمره‌ای بین ۰ تا ۱۰۰ لحاظ می‌شود

به علاوه دو ستون یکی شامل شناسه فرد و دیگری شامل دانشگاه محل تحصیل فرد است

از شما خواسته شده با توجه به دیگر اطلاعات، نمره پروژه را تخمین بزنید

ب) معرفی داده های سوال

اطلاعات	نام ستون
شناسه هر فرد	ID
دانشگاه محل تحصیل *	uni
نمره آزمون ساختمان داده	DS
نمره آزمون طراحی الگوریتم	ALG
نمره آزمون یادگیری ماشین	ML
نمره آزمون یادگیری عمیق	DL
نمره آزمون پایگاه داده	DB
نمره آزمون سخت افزار	HW
نمره آزمون سیستم عامل	OS
نمره آزمون شبکه های کامپیوتری	NTW
نمره آزمون پردازش سیگنال دیجیتال	DSP
نمره آزمون آمار و جبر خطی	STandLAL
نمره پروژه	Prj

*برخی دانشگاه هایی که در این مجموعه داده وجود دارد یک بار با نام کامل و یک بار با سرنام معرفی شده اند:

☐ دانشگاه صنعتی شریف که با Sharif University of Technology و SUT معرفی شده است.

☐ دانشگاه صنعتی اصفهان که با Isfahan University of Technology و IUT معرفی شده است.

☐ دانشگاه تهران که با UT , University of Tehran معرفی شده است.

در ضمن در این ستون ممکن است گاهی حروف کوچک و بزرگ برای یک نام وجود داشته باشد مثلاً isfahan university of technology و Isfahan University of Technology هر دو بیانگر یک دانشگاه است.

*داده این مجموعه از تمرین در لینک زیر موجود است:

https://drive.google.com/file/d/1JVLzDopXWjh_ObMvhMQxQLrECOkU6VWZ/view?usp=share_link

ج) بیان خواسته ها

در ابتدا باید مراحل تحلیل داده و پیش پردازش داده انجام شود، سپس بر پایه mlp و یکی دیگر از الگوریتم های معرفی شده در درس (مثل رگرسیون) مدلی برای تخمین نمره پروژه ارائه کنید

*رعایت نکات زیر الزامی است:

☐ تقسیم داده به train و test با نسبت ۰.۲ و random state برابر ۹۸

☐ استفاده از ماتریس درهم ریختگی و معیار accuracy, precision, recall

☐ تنظیم الگوریتم و شبکه عصبی به نحوی که بیش برازش و کم برازش رخ ندهد

*فایل کدها حتما باید به زبان پایتون و در ژوپیترونوت بوک باشد و به جز کتابخانه های زیر استفاده از بقیه کتابخانه ها ممنوع است:

Numpy, pandas, matplotlib, seaborn, plotly, sklearn, pytorch ,

tensorflow , keras, tqdm

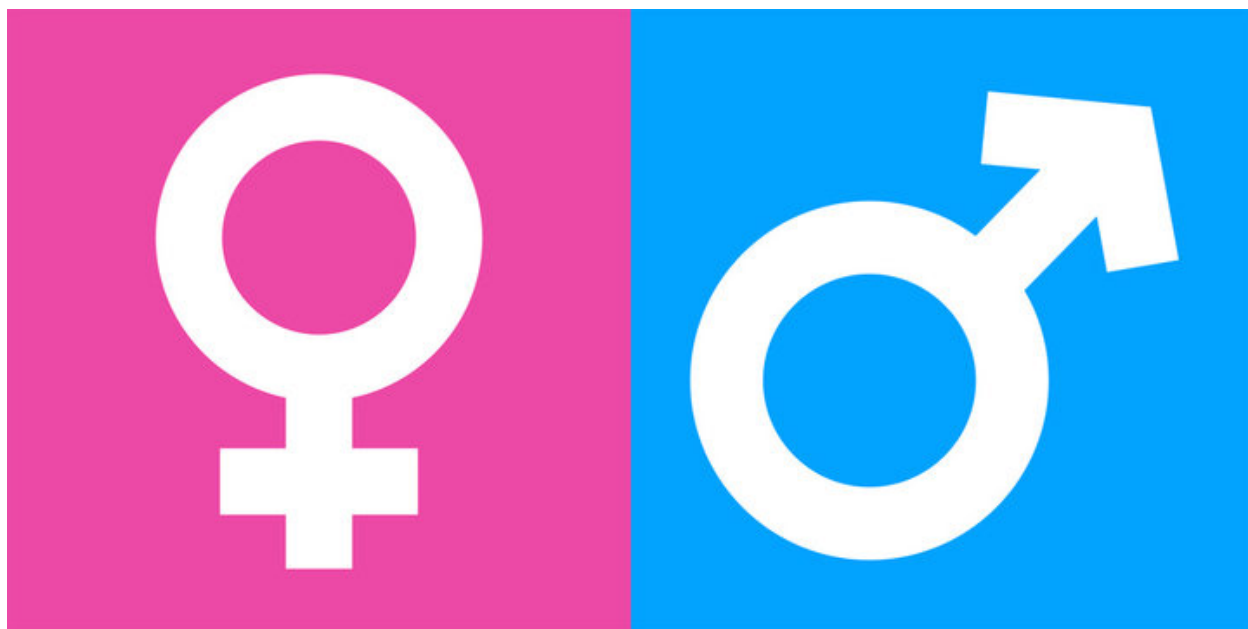
*در Markdown های نوت بوک علت هر گام در تحلیل داده و نتیجه دریافتی، همچنین شرح ماتریس در هم ریختگی و معیارهای ذکر شده باید به دقت بیان شود

همچنین بیان مسیر رسیدن به پارامترها و ساختار mlp و دیگر الگوریتم یادگیری ماشین که نتیجه مطلوب را بدهند لازم است.

(چنان چه بیان این نکات در Markdown دشوار است، در فایل pdf ای همراه تمرین ارسال شود)

سوال شماره ۲

جنسیت



الف) معرفی کلی مسئله

داده های زیر بر اساس برخی رفتارهای روزانه جمع آوری شده، ابتدا ستون جنسیت را نادیده بگیرید و داده ها را به دو خوشه خوشه بندی کنید. سپس با توجه به خوشه بندی انجام شده و با توجه به داده های ستون جنسیت با رسم نمودارهای مناسب به این پرسش پاسخ دهید که آیا خوشه بندی صورت گرفته با طبقه بندی جنسیتی داده ها مطابق است یا نه؛ همچنین بر اساس معیارهای silhouette score و davies bouldin score خوشه بندی صورت گرفته را ارزیابی کنید.

سپس بر پایه متد بازو برای الگوریتم Kmeans تعداد خوشه های مناسب داده ها را بیابید، در صورتی که عددی غیر از دو است، مجدداً بر اساس دو معیار گفته شده خوشه بندی را ارزیابی کنید

ب) معرفی داده های سوال

هر سطر بیانگر نظرسنجی صورت گرفته از یک فرد است.

اطلاعات	نام ستون
بیانگر رنگ مورد علاقه فرد است	Favorite Color
بیانگر سبک موسیقی مورد علاقه فرد است	Favorite Music Genre
بیانگر نوشیدنی مورد علاقه فرد است	Favorite Soft Drink
بیانگر جنسیت فرد است	Gender

***داده این مجموعه از تمرین در لینک زیر موجود است:

https://drive.google.com/file/d/1JVLzDopXWjh_ObMvhMQxQLrECOkU6VWZ/view?usp=share_link

ج) بیان خواسته ها

در ابتدا باید مراحل تحلیل داده و پیش پردازش داده انجام شود، سپس بر پایه الگوریتم K Mean مدلی برای خوشه بندی ارائه کنید

✳️ فایل کدها حتما باید به زبان پایتون و در ژوپیتر نوت بوک باشد و به جز کتابخانه های زیر استفاده از بقیه کتابخانه ها ممنوع است:

Numpy, pandas, matplotlib, seaborn, plotly, sklearn

✳️ در Markdown های نوت بوک علت هر گام در تحلیل داده و نتیجه دریافتی، همچنین شرح معیارهای ذکر شده و نتیجه آنان باید به دقت بیان شود

(چنان چه بیان این نکات در Markdown دشوار است، در فایل pdf ای همراه تمرین ارسال شود)

سوال شماره ۳

موسیقی ایرانی



الف) معرفی کلی مسئله

داده‌های زیر مربوط به برخی از موزیک‌های فارسی موجود در اسپاتیفای است، با توجه به داده‌ها

مدلی مبتنی بر الگوریتم برای پیش بینی انرژی ارائه دهید

ب) معرفی داده‌های سوال

هر سطر بیانگر ویژگی‌های یک قطعه موسیقی است.

اطلاعات	نام ستون
شناسه قطعه در اسپاتیفای	track_id
در برخی آلبوم‌ها (خصوصاً آلبوم‌های قدیمی) آلبوم در بیش از یک CD عرضه می‌شد، در این ویژگی شماره ای CD که قطعه مورد نظر در آن است، نوشته شده است	disc_number
دیوریشن مقدار زمان یک نت است.	duration_ms
ماهیت خشونت داشتن قطعه را بیان می‌کند	explicit
بیانگر محبوبیت قطعه است	popularity
شماره قطعه در آلبوم	track_number
شناسه آلبوم در اسپاتیفای	album_id
تعداد کل قطعاتی که در آلبوم است	album_total_tracks
تاریخ انتشار آلبوم	album_release_date
سال انتشار آلبوم	album_release_year
API ترک	track_href
API آلبوم	album_href
مقداری است عددی که به قطعات نسبت داده	danceability

	میشود
energy	نشان دهنده یک معیار از شدت قطعه است.
loudness	بلندی صدا
mode	مد قطعه مورد نظر
Speechiness	معیاری است موسیقایی است برای بیان لغات در قطعه
acousticness	آکوستیک بودن (معیار موسیقایی)
instrumentalness	این مقدار نشان دهنده میزان آواز در آهنگ است. هر چه به 1.0 نزدیکتر باشد،
liveness	این مقدار احتمال ضبط آهنگ توسط مخاطب را توصیف می کند
valence	معیاری است موسیقایی است برای بیان خوشحال کننده بودن یا غم انگیز بودن آهنگ است
tempo	ضربانگ قطعه

***داده این مجموعه از تمرین در لینک زیر موجود است:

https://drive.google.com/file/d/1JVLzDopXWJh_ObMvhMQxQLrECOkU6VWZ/view?usp=share_link

ج) بیان خواسته ها

در ابتدا باید مراحل تحلیل داده و پیش پردازش داده انجام شود، سپس بر پایه mlp و یکی دیگر از الگوریتم های معرفی شده (مثل رگرسیون) در درس مدلی برای تخمین انرژی ارائه کنید

*رعایت نکات زیر الزامی است:

- ☐ تقسیم داده به train و test با نسبت ۰.۲ و random state برابر ۹۸
- ☐ استفاده از ماتریس درهم ریختگی و معیار accuracy, precision, recall
- ☐ تنظیم الگوریتم و شبکه عصبی به نحوی که بیش برازش و کم برازش رخ ندهد

*فایل کدها حتما باید به زبان پایتون و در ژوپیتر نوت بوک باشد و به جز کتابخانه های زیر استفاده از بقیه کتابخانه ها ممنوع است:

Numpy, pandas, matplotlib, seaborn, plotly, sklearn, pytorch ,
tensorflow , keras, tqdm

*در Markdown های نوت بوک علت هر گام در تحلیل داده و نتیجه دریافتی، همچنین شرح ماتریس درهم ریختگی و معیارهای ذکر شده باید به دقت بیان شود

همچنین بیان مسیر رسیدن به پارامترها و ساختار mlp و دیگر الگوریتم یادگیری ماشین که نتیجه مطلوب را بدهند لازم است.

(چنان چه بیان این نکات در Markdown دشوار است، در فایل pdf ای همراه تمرین ارسال شود)

سوال شماره ۴

بیمه مسافرتی



الف) معرفی کلی مسئله

یک شرکت بیمه، بر اساس اطلاعاتی که از که دارد، از شما میخواهد مدلی برای این که شخص بیمه سفر را خریداری می کند یا نه را طراحی کنید

ب) معرفی داده های سوال

اطلاعات	نام ستون
سن هر مشتری	Age

Employment Type	بخشی که در آن مشتری مشغول به کار است
GraduateOrNot	آیا مشتری فارغ التحصیل دانشگاه است؟
AnnualIncome	درآمد سالانه مشتری
FamilyMembers	تعداد اعضای خانواده مشتری
ChronicDiseases	آیا مشتری از هر بیماری مهم یا شریطی مانند دیابت، فشار خون بالا یا غیره رنج می برد؟
FrequentFlyer	آیا مشتری در طول دو سال گذشته، حداقل چهار سفر داشته است؟
EverTravelledAbroad	آیا مشتری تا به حال به یک کشور خارجی سفر کرده است؟
TravelInsurance	آیا مشتری بسته بیمه مسافرتی، خریداری کرده است؟

***داده این مجموعه از تمرین در لینک زیر موجود است:

https://drive.google.com/file/d/1JVLzDopXWJh_ObMvhMQxQLrECOkU6VWZ/view?usp=share_link

ج) بیان خواسته ها

در ابتدا باید مراحل تحلیل داده و پیش پردازش داده انجام شود، سپس بر پایه mlp و یکی دیگر از الگوریتم های معرفی شده (مثل رگرسیون لجستیک) در درس مدلی برای تخمین این که بیمه سفر گرفته شده یا نه ارائه کنید

*رعایت نکات زیر الزامی است:

- ☐ تقسیم داده به train و test با نسبت ۰.۲ و random state برابر ۹۸
- ☐ استفاده از ماتریس درهم ریختگی و معیار accuracy, precision, recall
- ☐ تنظیم الگوریتم و شبکه عصبی به نحوی که بیش برازش و کم برازش رخ ندهد

*فایل کدها حتما باید به زبان پایتون و در ژوپیترونوت بوک باشد و به جز کتابخانه های زیر استفاده از بقیه کتابخانه ها ممنوع است:

Numpy, pandas, matplotlib, seaborn, plotly, sklearn, pytorch ,
tensorflow , keras, tqdm

*در Markdown های نوت بوک علت هر گام در تحلیل داده و نتیجه دریافتی، همچنین شرح ماتریس درهم ریختگی و معیارهای ذکر شده باید به دقت بیان شود

همچنین بیان مسیر رسیدن به پارامترها و ساختار mlp و دیگر الگوریتم یادگیری ماشین که نتیجه مطلوب را بدهند لازم است.

(چنان چه بیان این نکات در Markdown دشوار است، در فایل pdf ای همراه تمرین ارسال شود)