



دانشکده‌گان ریاضی، آمار و علوم کامپیوتر

درس: یادگیری ماشین آماری

استاد درس: دکتر امینی

پروژه درس یادگیری ماشین آماری

مهلت تحویل: تا پایان روز ۱۰م بهمن ماه

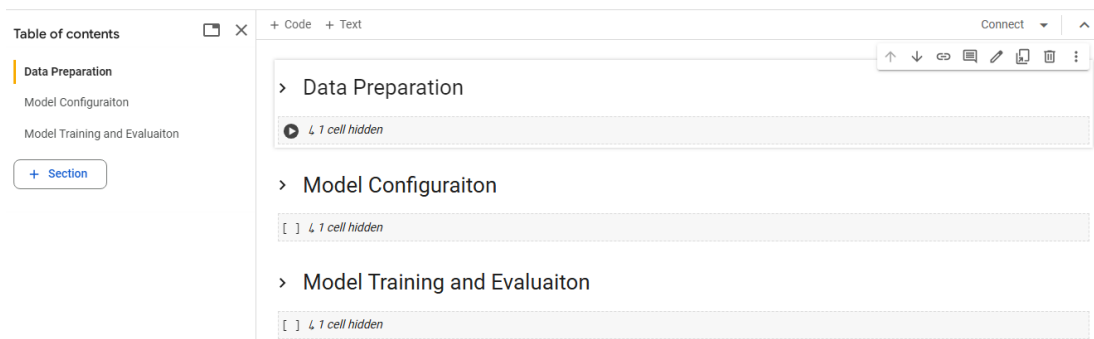
محل تحویل: سامانه ایلرن

لطفا در انجام پروژه به موارد زیر دقت کنید:

- نمره پروژه از ۱۰۰ و طبق بارمبندی زیر خواهد بود:
 - شرح مبسوط و توضیحات کامل و جامع مساله پروژه با فرمول، شکل و مثال (۴۰ نمره)
 - کد پایتون پیاده سازی مدل انتخاب شده (۳۰ نمره)
 - انتخاب و استفاده از دادگان مناسب مساله در کد (۲۰ نمره)
 - تحلیل و بررسی دقت مدل و نتایج بدست آمده (۱۰ نمره)
 - فایل پروژه بایستی به صورت یک فایل فشرده در سامانه ایلرن در قسمت "پروژه" بارگذاری گردد. محتوای فایل فشرده باید شامل موارد زیر باشد:
 - فایل pdf گزارش کار با فرمت توضیح داده شده در قسمت (توضیحات گزارش کار)
 - فایل نوتبوک کد در زبان پایتون شامل کد و نتایج اجرا با فرمت توضیح داده شده در قسمت (توضیحات پیاده سازی)
 - فایل مقاله یا مقاله‌های که در انجام پروژه از آنها استفاده کرده اید.
-

توضیحات پیاده سازی:

فایل نوتبوک پروژه را به سه بخش Data Preparation، Model Configuration و Model Training and Evaluation تقسیم کنید (مشابه تصویر زیر) و طبق توضیحات داده شده در ادامه، کدهای پروژه را به این سه بخش اضافه کنید. (در هر بخش به تعداد دلخواه می‌توانید سلول اضافه کنید. لزومی به قرار دادن همه کد در یک سلول نیست.)



بخش Data Preparation:

- در این قسمت کدهای مربوط به خواندن مجموعه داده انتخاب شده (و یا در صورتی که از مجموعه داده شبیه سازی شده استفاده می‌کنید، کدهای استفاده شده جهت تولید داده ها) و عملیات پیش پردازشی انجام شده روی مجموعه دادگان (مانند تقسیم مجموعه دادگان به مجموعه‌های آزمایشی و اعتبارسنجی، نرمال‌سازی، تبدیل داده های کیفی به کمی (در صورت وجود) و ...) را قرار دهید.
- تذکر مهم:** با توجه به اینکه تمام کدهای ارسال شده اجرا و با نتیجه ارسال شده مقایسه خواهد شد، در صورتی که در این بخش مولفه تصادفی وجود دارد که بر دقت نتایج مستقیماً تأثیر گذار است، مانند روش تقسیم داده ها به مجموعه‌های آموزشی و اعتبار سنجی و یا روش تولید داده های شبیه سازی شده، حتماً از روش تنظیم seed کنید تا امکان باز تولید نتایج بدست آمده وجود داشته باشد. مثلاً اگر از تابع `train_test_split` از کتابخانه `sklearn` برای تقسیم مجموعه دادگان به مجموعه‌های آموزشی و اعتبارسنجی استفاده می‌کنید، پارامتر `random_state` این تابع را مقداردهی کنید.

بخش Model Configuration

- در این بخش کدهای مربوط به پیاده‌سازی مدل خود، مقداردهی ابرپارامترها... را اضافه کنید. مثلاً اگر از شبکه های عصبی استفاده می‌کنید، کدهای مربوط به طراحی مدل، انتخاب ابرپارامترها، انتخاب `loss function`، `optimizer` و ... را اضافه کنید.
- تذکر مهم:** با توجه به اینکه تمام کدهای ارسال شده اجرا و با نتیجه ارسال شده مقایسه خواهد شد، در صورتی که در این بخش نیز مولفه تصادفی وجود دارد که بر دقت نتایج مستقیماً تأثیر گذار است، مانند مقدار دهی اولیه پارامترها (وزن‌ها)ی شبکه، حتماً از روش تنظیم seed کنید تا امکان باز تولید نتایج وجود داشته باشد. به عنوان مثال، در صورت استفاده از پایتورچ، از تابع `torch.manual_seed` استفاده کنید.

بخش Model Training and Evaluation:

- در این قسمت کدهای مربوط به آموزش، ارزیابی و گزارش نتایج را اضافه کنید.

توضیحات گزارش کار

گزارش پروژه باید شامل بخش‌های **چکیده** (در حدود یک پاراگراف به صورت خلاصه هدف پروژه، دیتاست استفاده شده و نتایج بدست آمده را توضیح دهید.)، **مدل** (توضیحات کامل، جامع و مبسوط مدل انتخاب شده)، **مجموعه دادگان** (معرفی کامل مجموعه دادگان، دلیل اینکه چرا این مجموعه دادگان را برای مدل خود مناسب دانسته و انتخاب کرده‌اید)، **توضیحات پیاده‌سازی** (توضیح خط به خط کد لازم نیست فقط بخش‌های اصلی کد را توضیح دهید. در صورت استفاده و ایده گرفتن از کدهای موجود در گیت هاب، kaggle، کتاب‌ها و... لطفا منبع کد را در بخش منابع اضافه کنید.)، **تحلیل نتایج** و **منابع** (در صورتی که از چندین مرجع استفاده کرده اید) باشد.

با آرزوی موفقیت