



دانشگاه صنعتی اصفهان
دانشکده مهندسی برق و کامپیوتر

درس مبانی یادگیری ماشین

تکلیف عملی دوم

مهلت تحویل: ۱ دی ۱۴۰۲

در این تکلیف شما مجاز به استفاده از کتابخانه‌هایی مانند scikit-learn هستید، به غیر از سولاتی که در صورت آن ذکر شده است که باید از پایه پیاده‌سازی کنید.

سوال ۱

در این سوال قصد داریم که یک طبقه بند دو کلاسه را طراحی کنیم. این یک مجموعه داده تصمیم خرید خودرو است که نشان می‌دهد آیا مشتری خودرو خریده است یا خیر. این مجموعه داده شامل جزئیات ۱۰۰۰ مشتری با توجه به حقوق سالانه آنها است که قصد خرید ماشین را دارند. ستون‌های این دیتاست به شرح زیر است:

- Gender: جنسیت
- Age: سن
- AnnualSalary: درآمد سالانه
- Purchased: این ستون شامل برچسب‌های داده است، که اگر یک باشد، ماشین خریداری شده و در غیر اینصورت ماشین خریداری نشده است.

در شکل زیر چند سطر از دیتاست این سوال قابل مشاهده است.

Gender	Age	AnnualSalary	Purchased
Male	35	20000	0
Male	40	43500	0
Male	49	74000	0
Male	40	107500	1

الف) ابتدا با استفاده از کتابخانه‌های مناسب داده ورودی را از فایل car_data.csv بخوانید، و ۵ سطر اول آن را نمایش دهید. (۲ نمره)

ب) نمودار مقادیر ستون‌های سن و جنسیت را به ترتیب در قالب Histogram و Pie Chart نشان دهید. (۳ نمره)

ج) باتوجه به اینکه تمامی داده‌ها در دسته‌بند باید به صورت عددی باشند، یکسری از ویژگی‌ها به صورت غیر عددی هستند آن‌ها را شناسایی کنید و تبدیل به مقادیر عددی متناظر کنید. (۱۰ نمره)

ه) ابتدا داده‌ها را با استفاده از روش StandardScaler نرمال‌سازی کنید، سپس ۸۰ درصد داده‌ها را به عنوان داده آموزشی و ۲۰ درصد آن را برای تست قرار دهید. (۱۰ نمره)

ک) تابعی بنویسید که داده‌های آموزشی را به عنوان ورودی دریافت کند و مدل Logistic Regression را از پایه بر روی این داده‌ها اجرا کند و در هر گام وزن‌های مورد نظر را به روزرسانی کند. (۳۰ نمره)

نکته: تابع Cross-Entropy را به عنوان خطا در نظر بگیرید.

خ) نمودار خطا و دقت بر روی داده‌های آموزشی را بر اساس گام‌های بروز رسانی وزن‌ها رسم کنید. (۵ نمره)

م) میزان خطا و دقت مدل آموزش دیده را بر روی داده‌های تست را گزارش کنید. (۱۰ نمره)

سوال ۲

در این سوال قصد داریم با استفاده از مدل طبقه‌بند بردار پشتیبان^۱ جنسیت گربه‌ها را با داشتن اطلاعات زیستی آن‌ها طبقه‌بندی کنیم. اطلاعات این دیتاست به شرح زیر است:

• Gender: جنسیت

• Bwt: وزن بدن

• Hwt: وزن قلب

در جدول زیر چند سطر از این مجموعه داده قابل مشاهده است.

Gender	Bwt	Hwt
F	2	7
F	2	7.4
M	2.8	10.2
M	2.9	11.3

الف) ابتدا داده‌ها از فایل cats.csv بخوانید، و همه مقادیر آن را به مقادیر عددی تبدیل کنید، سپس ۸۰ درصد داده‌ها را برای آموزش و باقی آن‌ها را برای تست در نظر بگیرید. (۲ نمره)

ب) داده‌ها را با استفاده از روش StandardScaler نرمال کنید و آن را بر روی یک صفحه دو بعدی یک بار قبل از نرمال‌سازی و یک بار بعد از نرمال‌سازی نشان دهید، به صورتی که محور افقی مقادیر Hwt و محور عمودی مقادیر Bwt باشد. مشاهدات خود را گزارش کنید. (۵ نمره)

ج) با استفاده از تابع `np.logspace(start, stop, num)` آرایه‌های زیر را ایجاد کنید: (۵ نمره)

• آرایه C که شامل ۸ عدد در بازه ۲- و ۵

• آرایه Gamma که شامل ۱۶ عدد در بازه ۶- و ۱

د) با استفاده از تابع SVC در کتابخانه scikit-learn سعی کنید یک طبقه‌بند بردار پشتیبان خطی تهیه کنید، و همچنین بهترین مقادیر C و Gamma را با استفاده از روش Grid Search گزارش کنید. (۱۰ نمره)

خ) عملیات سوال قبل را با استفاده از کرنل گوسی اجرا کنید. با در نظر گرفتن دقت این دو مدل بر روی داده‌های آموزشی کدام مدل عملکرد بهتری دارد؟ چرا؟ (۵ نمره)

م) حال با استفاده از بهترین مدل ساخته شده در قسمت قبل، عملکرد این مدل را بر روی داده‌های تست بیابید. (۳ نمره)

^۱SVM

نکات تکمیلی

۱. برای انجام این تکلیف استفاده از زبان پایتون الزامی است.
۲. تکالیف را در محیط jupyter notebook پیاده‌سازی کنید و فایل ipynb را ارسال کنید.
۳. توضیح کدی که نوشته‌اید، بررسی و تحلیل نتایج آن و بیان علت نتایج و نیز مقایسه نتیجه با آنچه مورد انتظارتان بوده است، از اهمیت بالایی برخوردار است. شما می‌توانید گزارش پروژه را در همان محیط jupyter notebook بنویسید و نیازی به فایل pdf جداگانه نیست. هم‌چنین اگر برای حل سوال فرضیات خاصی مدنظر دارید حتماً آن را در متن گزارش قید کنید.
۴. فرمت نامگذاری تکلیف ارسالی باید به صورت زیر باشد: HWX_Programming_LastName_StudentID که X شماره تکلیف LastName نام خانوادگی شما و StudentID شماره دانشجویی شما است.
۵. انجام این تکلیف به صورت تک نفره است. در صورت مشاهده تقلب، نمرات هم مبدا کپی و هم مقصد آن صفر لحاظ می‌شود.
۶. شما می‌توانید تا یک هفته پس از پایان مهلت تکلیف آن را در یکتا بارگذاری کنید. در این صورت به ازای هر روز تاخیر ۵ درصد از نمره تکلیف کسر می‌شود. پس از اتمام این یک هفته امکان ارسال با تاخیر وجود ندارد.
۷. در صورت وجود هر گونه ابهام و یا سوال می‌توانید سوالات خود را در گروه تلگرام بپرسید. هم‌چنین می‌توانید برای رفع ابهامات با دستیاران آموزشی از طریق تلگرام در تماس باشید.

آیدی‌ها:

Matin_Fazel

موفق باشید.