



تمرین سری دوم درس مباحث ویژه

نام مدرس: دکتر محمدی

دستیار آموزشی مرتبط: امیررضا فاتح

مهلت تحویل: ۱۷ مهر ۱۴۰۰

۱- در این تمرین می‌خواهیم بدون پیاده‌سازی به صورت عملی، با استفاده از جدول زیر تابعی بیابیم که پیش‌بینی کند هر فرد آیا توان پرداخت وام را دارد یا خیر. برای اینکار از Logistic Regression با تابع فعال‌ساز sigmoid بهره می‌گیریم. فرض کنید که وزن‌های اولیه و بایاس برابر با ۱ و تابع ضرر log loss باشد.

ا) پس از محاسبه loss بر مبنای وزن‌های اولیه، وزن‌ها را بر مبنای گرادیان کاهشی تصادفی (Stochastic Gradient Descent) به روز نمایید. (نحوه انتخاب داده‌ها را بر مبنای جدول از بالا به پایین در نظر بگیرید) این عملیات را دوبار تکرار، نتایج را گزارش و مراحل محاسبه را یادداشت نمایید.

$$w_1 = 1$$

$$w_2 = 1$$

$$b = 1$$

$$\text{sigmoid}(z) = \frac{1}{1 + e^{-z}}$$

$$\log \text{ loss} = -\frac{1}{N} \sum_{i=1}^N (y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}))$$

$$\text{Learning rate} = 0.05$$

$$\text{batch size} = 2$$

توان پرداخت (y)	داشتن کار (x ₂)	سن (x ₁)
ندارد	دارد	۲۲
ندارد	ندارد	۲۵
دارد	دارد	۴۷
ندارد	ندارد	۵۲
دارد	دارد	۴۶
دارد	دارد	۵۶
ندارد	ندارد	۵۵
دارد	ندارد	۶۰

ب) دو روش گرادیان کاهشی و گرادیان کاهشی تصادفی را با هم مقایسه و مشکلات هر یک را ذکر کنید. چه راهکاری برای بهبود عملکرد آنها ارائه می‌دهید؟ (در صورت استفاده از منبع، حتما ذکر شود)

۲- مجموعه داده زیر را در نظر بگیرید که مربوط به یک مسئله باینری است و ورودی آن یک‌بعدی است. با استفاده از دو روش Linear Regression و Logistic Regression مرز تصمیم میان دو کلاس را یک بار محاسبه کنید (در این مثال مرز تصمیم xی است که خروجی مدل برابر با ۰.۵ باشد). نتیجه بدست آمده را به طور دقیق تحلیل کنید و دو روش را مقایسه کنید.

n	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
x	۲.۳	۱.۴	۲.۶	۳.۱	۱.۸	۲.۸	۵.۴	۶.۳	۵.۸	۶.۷	۴.۹	۴۵.۲
y	۰	۰	۰	۰	۰	۰	۱	۱	۱	۱	۱	۱



تمرین سری دوم درس مباحث ویژه

نام مدرس: دکتر محمدی
دستیار آموزشی مرتبط: امیررضا فاتح

مهلت تحویل: ۱۷ مهر ۱۴۰۰

۳- در این تمرین می‌خواهیم با استفاده از Logistic Regression انواع زنبق‌ها را از هم تشخیص دهیم. برای این کار از مجموعه داده iris و تنها از دو ویژگی اول آن استفاده می‌کنیم. در این تمرین از ماژول LogisticRegression در کتابخانه scikit-learn استفاده نمایید. پس از جداسازی ۳۰٪ داده‌ها برای فاز آزمون، مراحل آموزش و آزمون را کامل نمایید.

دقت داشته باشید که تعداد کلاس‌ها بیشتر از دو کلاس است لذا در صورت نیاز تغییرات لازم را در پارامترهای ماژول LogisticRegression اعمال کنید.

ضمن انجام آزمایشات، گزارشی بنویسید که در آن سوالات زیر پاسخ داده شده‌اند.

در گزارش خود:

- ا) ابتدا مجموعه داده iris را معرفی کنید.
- ب) نمای دوبعدی داده‌های آموزش را طوری رسم کنید که کلاسها متمایز باشند.
- ج) داده‌های آزمون را نیز به همان ترتیب بخش قبل نمایش دهید.
- د) دقت فاز آموزش و آزمون چقدر است؟ تحلیل خود را از مقادیر بدست آمده بنویسید.
- ه) در نهایت Confusion matrix را کشیده و تحلیل‌های خود را از آن بنویسید.

نکات تکمیلی:

۱. لطفاً پاسخ سوالات (تئوری و توضیحات پیاده‌سازی) را به طور گویا و به زبان فارسی و در صورت امکان تایپ همراه با سورس کدهای نوشته شده، در یک فایل فشرده شده به شکل HW2_YourStudentID.zip قرار داده و بارگذاری نمایید.
۲. منابع استفاده شده را به طور دقیق ذکر کنید.
۳. برای سهولت در پیاده‌سازی‌ها و منابع بیشتر، زبان پایتون پیشنهاد می‌شود. لطفاً کدهای مربوطه را در فرمت ipynb ارسال نمایید و هر کدام از موارد خواسته شده در بالا را در یک سلول جدید پیاده‌سازی نمایید.
۴. ارزیابی تمرین‌ها براساس صحیح بودن راه حل‌ها، گزارش مناسب، بهینه بودن کدها و کپی نبودن می‌باشد.
۵. در مجموع تمام تمرین‌ها، تنها ۷۲ ساعت تاخیر در ارسال پاسخ‌ها مجاز است اما پس از آن به صورت خطی از نمره شما کسر خواهد شد (معادل با روزی ۵۰ درصد).
۶. تمرین‌ها باید به صورت انفرادی انجام شوند و حل گروهی تمرین مجاز نیست.
۷. پرسش و پاسخ در رابطه با تمرین‌ها را می‌توانید در گروه مربوطه مطرح کنید.

موفق و سربلند باشید