



## درس یادگیری ماشین

مدرس: دکتر سامان هراتی زاده

الگوریتم‌های Linear R, Logistic R

نیم سال اول ۹۶-۱۳۹۵

تمرین شماره‌ی سه

تاریخ تحویل حضوری: ۱۳۹۵/۹/۷

مهلت ارسال تمرین: ۱۳۹۵/۹/۶

### تایید اعتبار اسکناس (۱۰۰ نمره)

(۱) هدف از این تمرین، پیاده‌سازی برنامه‌ای برای دسته‌بندی داده‌ها بر مبنای الگوریتم Logistic Regression است.

#### داده‌ها

داده پیوست شده مربوط به تأیید اعتبار (جعلی یا اصلی بودن) است. این داده‌ها دارای ۵ مولفه (۴ ویژگی و برچسب کلاس) و ۱۳۷۲ نمونه است. تمام ۴ ویژگی عددی هستند. مقدار ۱ برای برچسب کلاس نشان‌دهنده‌ی جعلی بودن اسکناس و مقدار ۰ نشان‌دهنده‌ی اصلی بودن اسکناس است.

#### ۱-۱) پیاده‌سازی الگوریتم (۷۰ نمره)

پیاده‌سازی مطلوب باید قاعده‌ی زیر را رعایت کند:

۱-۱-۱) با استفاده از الگوریتم Logistic Regression مدلی بسازید و برچسب کلاس را پیش‌بینی کنید. میزان دقت خود را روی

داده آزمون با استفاده از معیارهای  $accuracy$ ،  $precision$ ،  $recall$  و  $F1-Score$  گزارش کنید. این مراحل را یکبار

باروش  $5-fold-crossvalidation$  و یکبار با روش  $5-fold-feedforward-crossvalidation$  انجام دهید و معیارها

را در هر مرحله و میانگین کل آن‌ها را برای هر یک از روش‌ها گزارش کنید.

#### ۱-۲) تحلیل پارامتر (۳۰ نمره)

۱-۲-۱) با تغییر ضریب یادگیری در بازه‌ی  $[0, 1]$  تأثیر ضریب یادگیری را در سرعت همگرایی بررسی کنید. برای اینکار نموداری

بکشید که تعداد تکرارهای الگوریتم را بر حسب ضریب یادگیری نمایش دهد (دوستان جان توجه کنید در این حالت باید

شرط توقف را یک دلتای ثابت برای همه‌ی ضرایب یادگیری مختلف در نظر بگیرید. هم‌چنین برای تعداد معقولی از ضرایب بررسی فوق را حاصل نمایید (انتخاب معقول بودن این تعداد بر عهده‌ی شماست)

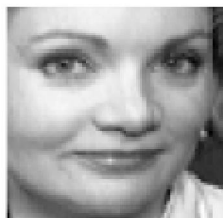
۱-۲-۲) تاثیر ضریب یادگیری را روی معیارهای *accuracy*، *precision* و *recall* بررسی کنید. برای اینکار برای هر یک از معیارها یک نمودار بکشید و میانگین آن را بر روی داده‌ی آزمون بر حسب ضریب یادگیری نمایش دهید (دوستانِ جانِ باز هم توجه کنید در این حالت نیز باید شرط توقف را یک دلتای ثابت برای همه‌ی ضرایب یادگیری مختلف در نظر بگیرید).

۱-۲-۳) به نظر شما کاربرد روش *k-fold-feedforward-crossvalidation* در چه نوع داده‌هایی است؟ چرا؟

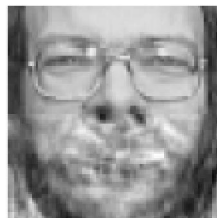
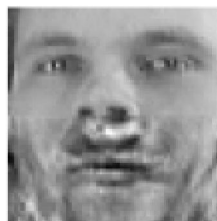
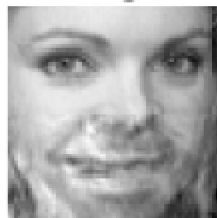
## پیش‌بینی نیم‌چهره (۲۰۰ نمره)

۲) هدف از این تمرین استفاده از الگوریتم multi-output Linear Regression برای پیش‌بینی نیمه‌ی پایینی چهره‌ی فرد با در اختیار داشتن نیمه‌ی بالایی چهره‌ی اوست (مانند شکل زیر).

true faces



Linear regression



## داده ها

داده‌ی پیوست شده شامل ۴۰۰ عکس سیاه سفید با ابعاد ۶۴ در ۶۴ از چهره‌های ۳۹ فرد مختلف است. (دوستان توجه کنید که فایل پیوست شده یک فایل با پسوند mat. است که قابل استفاده در نرم افزار متلب است. اگر محیط برنامه نویسی تان چیزی جز متلب است دو کار (شاید هم بیشتر) انجام دهید. اول اینکه فایل را با متلب باز کنید و سپس با یک پسوند استاندارد مانند CSV. آن را ذخیره کنید و در محیط برنامه نویسی خود بارگذاری کنید. هم چنین اگر در محیط برنامه نویسیتون پایتون هست با استفاد از پکیج `scipy.io.matload()` می‌توانید فایل مذکور را بارگذاری کنید.

### ۲-۱) پیاده سازی الگوریتم (۲۰۰ نمره)

پیاده سازی مطلوب باید قاعده‌های زیر را رعایت کند:

۲-۱-۱) برای ساخت مدل ماتریس مربوط به هر عکس را در راستای ارتفاع به دو قسمت مساوی تقسیم کنید به این ترتیب که پیکسل های مربوط به نیمه‌ی بالایی عکس بردار ویژگی را تشکیل دهند و پیکسل های مربوط به نیمه‌ی پایینی عکس بردار کلاس (بله بردار عکس و نه تنها یک واحد عددی!) را نشان می‌دهند. برای درک بهتر فرمالاسیون این الگوریتم یک فایل متنی با نام `multi-output-linear-regression` پیوست شده است. (لطفا حتما پیش از شروع به پیاده سازی این فایل را مطالعه کنید)

۲-۱-۲) از داده‌های مربوط به ۳۰ نفر برای آموزش و از داده‌های مربوط به ۸ نفر باقیمانده به عنوان آزمون استفاده کنید. معیار `RMSE` (`Root Mean Squared Error`) بر روی داده‌ی آزمون گزارش کنید. بهترین شکل تکمیل شده ی ۸ نفر موجود در داده‌ی آزمون را در گزارش رسم کنید.

توجه: پیاده سازی قسمت ۲ این تمرین با چالش‌های بیشتری نسبت با تمرین پیش روبه‌روست (شاید..)، از این رو نمره‌ی تخصیص داده‌شده به آن دو برابر سایر تمرین است و ارزیابی آن متناسب با نمره ی تخصیص داده‌شده به آن است.

### نکات مهم:

۱. نمره سوالات پیاده سازی و تحلیلی با کد به صورت زیر در نظر گرفته میشود:

۱. کدها ۴۰٪

۲. گزارش ۳۰٪

۳. تحویل حضوری ۳۰٪

۱۱. نمره سوالات تحلیلی بدون کد به صورت زیر در نظر گرفته میشود:

۱. گزارش ۶۰٪

۲.تحويل حضوری %۳۰

۱۱۱.نمرات کدها، گزارش و تحويل حضوری منوط به ارسال به موقع **کدها و گزارش** است