



دانشگاه تهران
پردیس دانشکده های فنی
دانشکده برق و کامپیوتر



پروژه نهایی یادگیری ماشین

دکتر اعرابی
دکتر ابوالقاسمی

آذر 1403

فهرست مطالب

فصل ۱	۳
مقدمه	۳
فصل ۲	۴
تعریف مسئله و شرح پروژه	۴
۲.۱ تعریف کلی مسئله	۴
۲.۲ دیتاست	۵
۲.۳ گروه‌بندی	۶
۲.۴ گزارش اولیه	۶
۲.۵ پیش پردازش داده و استخراج ویژگی	۹
۲.۶ طبقه‌بندی	۱۰
۲.۷ خوشه‌بندی	۱۲
فصل ۳	۱۳
گزارش کار	۱۳
فصل ۴	۱۴
بارم بندی	۱۴
فصل ۵	۱۵
نکات پایانی	۱۵

فصل ۱

مقدمه

یادگیری ماشین به عنوان یکی از ابزارهای کلیدی در تحلیل و درک الگوهای داده‌ها، در زمینه‌های مختلفی از جمله تشخیص و احراز هویت مورد استفاده قرار می‌گیرد. هدف این پروژه، آشنایی دانشجویان با مفاهیم پایه یادگیری ماشین از طریق کاربرد عملی آن در دو حوزه مرتبط با صدا است: تشخیص جنسیت گوینده و احراز هویت بر اساس ویژگی‌های صوتی.

در این پروژه، مدلی مبتنی بر یادگیری ماشین طراحی خواهید کرد که ابتدا قادر به پیش‌بینی جنسیت گوینده بر اساس ویژگی‌های صوتی در حوزه فرکانس یا زمان باشد و در ادامه بتواند هویت گوینده را با استفاده از این ویژگی‌های صوتی تشخیص دهد. داده‌های صوتی مورد استفاده در این پروژه شامل ویژگی‌های متنوعی هستند که بین افراد مختلف تفاوت‌های آشکاری دارند. برای موفقیت در این پروژه، لازم است مراحل مختلفی از جمله پیش‌پردازش داده‌ها برای حذف نویز، استخراج ویژگی‌های مفید صوتی با استفاده از روش‌های مطرح در یادگیری ماشین و پردازش سیگنال‌های صوتی، و طبقه‌بندی و خوشه‌بندی داده‌ها بر اساس این ویژگی‌ها را اجرا کنید. این فرآیند علاوه بر آشنایی با اصول یادگیری ماشین، دانش عملی دانشجویان را در برخورد با چالش‌های واقعی مانند انتخاب ویژگی مناسب و بهبود دقت مدل افزایش خواهد داد.

تعریف مسئله و شرح پروژه

۲.۱ تعریف کلی مسئله

امروزه آنالیز و پردازش سیگنال‌های گفتاری که شامل مطالعه و استخراج اطلاعات مفید از سیگنال‌های صوتی می‌شود کاربردهای گسترده‌ای یافته است. این فرایند شامل پردازش اولیه داده‌های صوتی و سپس استخراج ویژگی‌های متنوع از آنها است که بتوان با استفاده از الگوریتم‌های یادگیری ماشین در کاربردهای گوناگونی از جمله تشخیص گوینده صدا، تشخیص احساسات، فهم جملات و تبدیل آنها به متن و ... استفاده کرد. **سیگنال گفتاری شامل ویژگی‌های ارزشمندی در حوزه زمان و فرکانس است** که به ما این امکان را می‌دهد تا بتوانیم ابزارهای مختلفی را جهت پردازش آنها گسترش دهیم. در این پروژه با دو کاربرد تشخیص جنسیت گوینده‌ی صدا و همچنین تشخیص هویت آن آشنا خواهیم شد.

به طور کلی این پروژه در دو فاز تعریف می‌شود:

فاز اول: به طور کلی هدف از این فاز، آشنایی با سیگنال‌های صوتی به خصوص سیگنال‌های گفتاری و کاربردهای پردازشی آنها می‌باشد. در این فاز از شما انتظار می‌رود گزارش کوتاهی در مورد نوع داده‌های سیگنال‌های گفتاری و سپس چالش‌هایی که برای تشخیص الگو از این سیگنال‌ها وجود دارد، آماده کنید. در هر مرحله، لازم است پس از توضیح چالش مدنظر، راهکار(های) مرتبط با حل آن چالش را به صورت مختصر توضیح داده و مزایا و معایب آن را با ارجاع به منابع معتبر شرح دهید. برای مثال یکی از روش‌های مهم استخراج ویژگی از سیگنال‌های گفتاری استفاده از الگوریتم $MFCC^1$ است. ابتدا این الگوریتم‌ها را توضیح داده و سپس چالشی که هر روش برای

¹ Mel-frequency cepstral coefficients

حل آن مناسب است و مزایا و معایب آن در حل این چالش را معرفی کنید. در ادامه جزئیات بیشتری در مورد روند کار توضیح داده خواهد شد.

فاز دوم: در این فاز، به صورت عملی با استفاده از داده‌هایی که به صورت آماده در اختیار شما قرار داده خواهد شد، به پیاده‌سازی مراحل می‌پردازید. دیتای مورد نظر، سیگنال‌های گفتاری است که در طول ترم از شما بابت گزارش حل تمرین‌های درس جمع‌آوری شده است. در این فاز لازم است تا ابتدا داده‌ها را به صورت مناسب پیش‌پردازش کرده تا آن‌ها را به فرمت مناسب تبدیل کرده و همچنین نویز و یا تداخلات اضافی که ممکن است به مراحل بعد منتقل شود را حذف کنید. سپس با استفاده از روش‌های استخراج ویژگی که در ادامه به آن‌ها اشاره شده است یا با توجه به روش‌هایی که خودتان مدنظر دارید و در فاز قبل راجع به آن‌ها اطلاعاتی جمع‌آوری کرده‌اید، به استخراج ویژگی از داده‌ها بپردازید. در نهایت اگر ویژگی‌های استخراجی مناسب باشند، باید به طبقه‌بندی و خوشه‌بندی داده‌ها بپردازید. در صورت عدم موفقیت در این مرحله، ممکن است به ویژگی‌های بهتری احتیاج داشته باشید تا بتوانید به دقت مناسبی برسید.

۲.۲ دیتاست

پوشه‌ای که در اختیار شما قرار داده می‌شود شامل دیتاست فایل‌های صوتی به طول‌های متنوع و با فرمت زیر است:

HWx_Qy_Student-ID_gender.mp3

برای هر تسک از جمله تشخیص جنسیت و تشخیص هویت فرد گوینده، ابتدا باید دیتاست را بر اساس جنسیت یا شماره دانشجویی به دسته‌های مناسب تقسیم‌بندی کرده و سپس به آماده‌سازی و پیش‌پردازش آن برای استخراج ویژگی بپردازید. توجه کنید از آنجایی که داده‌ها دارای طول متغیر هستند، ابتدا لازم است طبق روشی مناسب داده‌ها را به پنجره‌های زمانی کوچکتر تقسیم کرده و از بخشی از آن‌ها در ادامه استفاده کنید ولی اجباری

به استفاده از طول پنجره مشخص یا روش مشخصی وجود ندارد. همچنین از آنجایی که فرمت فایل‌ها mp3 است، می‌توانید در صورت تمایل آن‌ها را به فرمت‌های دیگری که ممکن است برای شما مفید باشد تبدیل کنید. همچنین برای هر داده به نرخ نمونه‌برداری² آن توجه کنید و اطمینان حاصل کنید که تمام داده‌ها با نرخ یکسانی نمونه‌برداری شده باشند. با توجه به حجم بالای داده‌ها، لزومی به استفاده از تمام آن‌ها نیست که در ادامه در هر بخش بیشتر توضیح داده خواهد شد.

لینک داده‌ها: (داده‌های تمرین‌های دیگر به مرور به شما داده می‌شود)

https://drive.google.com/drive/folders/1pq_jGqdBda_QjNnK2yAzD4N2grbPF8Rs?usp=sharing

۲.۳ گروه‌بندی

شما می‌توانید به صورت انفرادی و یا گروه‌های حداکثر چهار نفره فعالیت داشته باشید. اسامی افراد گروه را یک نفر به نمایندگی **تا آخر ۲۰ آذر** در محل تحویل گروه‌بندی ارسال کند.

۲.۴ گزارش اولیه

در این مرحله لازم است به صورت گروهی گزارشی تهیه کنید.

گزارش اولیه با هدف آشنایی با مفهوم voice authentication، چالش‌های مرتبط با آن و روش‌های به کار گرفته شده در پیش‌پردازش و استخراج ویژگی‌های داده‌های صوتی است. همچنین در این گزارش شما با دو مفهوم open-set authentication و closed-set authentication به همراه چالش‌های پیاده‌سازی این دو روش آشنا

² Sampling rate

خواهید شد. هدف این است که درک عمیقی نسبت به پیشرفت ها و تکنیک های موجود در این زمینه حاصل گردد. همچنین مبحث طبقه بندی جنسیت (Gender Classification) به عنوان یکی از کاربردهای کلیدی مورد بررسی قرار خواهد گرفت.

این گزارش باید شامل موارد زیر باشد:

۱. مقدمه ای بر voice authentication

- تعریف voice authentication و اهمیت آن
- بررسی کاربرد های voice authentication مانند شناسایی گوینده و تشخیص جنسیت گوینده
- تعریف closed-set authentication و open-set authentication و تفاوت های این دو با هم
- بررسی چگونگی پیاده سازی این دو روش
- بررسی کاربرد های آن در voice authentication

۲. چالش های voice authentication

- شناسایی و توضیح چالش های اصلی که در تحقیقات و کاربردهای voice authentication و gender classification وجود دارد.

- بررسی راه حل های بالقوه و تحقیقات جاری برای غلبه بر این چالش ها.

۳. پیش پردازش داده های صوتی

- بحث در مورد اهمیت پیش پردازش داده های صوتی در زمینه voice authentication و

gender classification

- توضیح مراحل معمول پیش پردازش با تمرکز بر:

- Noise reduction (حذف نویز)

- Normalization (نرمالسازی داده ها)
- Windowing (تقسیم سیگنال به پنجره های زمانی)

۴. تکنیک های استخراج ویژگی

- توضیح تکنیک های مختلف استخراج ویژگی که در تجزیه و تحلیل داده های صوتی به کار می روند. ویژگی هایی مانند:

- MFCC (Mel Frequency Cepstral Coefficients)
- Fast Fourier Transform
- Log Mel Spectrogram
- Spectral Centroid
- Chroma Features
- Spectral Contrast
- Zero-Crossing Rate
- Linear Predictive Coding (LPC)
- Perceptual Linear Prediction (PLP)

۵. **Similarity Learning** (امتیازی)

- تعریف Similarity Learning در حوزه تحلیل صدا و نحوه استفاده از آن برای تشخیص شباهت بین ویژگی های صوتی.

- معرفی و توضیح Loss functions رایج در Similarity Learning (مانند: Contrastive Loss و Triplet Loss و)

۲.۵ پیش پردازش داده و استخراج ویژگی

در این بخش شما باید یاد بگیرید که چگونه داده‌های صوتی را آماده کنید و روش‌های مختلف استخراج ویژگی را برای آشنایی با ویژگی‌های سیگنال‌های صوتی به کار ببرید. لازم به ذکر است که قبل از شروع مراحل، تحلیل اولیه داده‌ها (مانند مشاهده طیف فرکانسی) می‌تواند در انتخاب تکنیک‌های مناسب برای پیش‌پردازش و استخراج ویژگی‌ها کمک‌کننده باشد.

این بخش شامل مراحل زیر است:

۱. بارگذاری داده‌ها

- فایل‌های صوتی خام را بارگذاری کنید.

۲. پیش‌پردازش داده‌های صوتی:

- در صورت وجود نویز، برای حذف آن می‌توان از تکنیک‌هایی مانند **Spectral Subtraction** یا

Bandpass Filtering استفاده کرد. بسته به نوع داده‌ها و نیاز پروژه، استفاده از این روش‌ها

ممکن است مفید باشد، اما انتخاب روش مناسب به تحلیل نویز و ویژگی‌های سیگنال بستگی دارد.

- اطمینان حاصل کنید که تمام داده‌ها با نرخ نمونه‌برداری یکسان پردازش شوند. در صورت لزوم از **resampling** استفاده کنید.

- برای تقسیم سیگنال به پنجره‌های زمانی کوچک‌تر و قابل تجزیه، از ابزارهایی مانند کتابخانه **librosa** یا **scipy.signal** استفاده کنید.

- سیگنال‌ها را نرمالسازی کنید تا داده‌ها یکنواخت و قابل مقایسه شوند.

۳. استخراج ویژگی‌های صوتی:

- ویژگی‌هایی مانند Spectral Centroid, Spectral MFCCT, Log Mel Spectrogram, Spectral Contrast و Bandwidth که اطلاعات فرکانسی را مدل می‌کند، استخراج کنید.
- ویژگی‌های زمانی مانند Zero-Crossing Rate و Energy برای شناسایی تغییرات سریع در صدا و بررسی تغییرات زمانی و شدت سیگنال استفاده شود.
- در صورت نیاز، از ترکیب ویژگی‌های زمانی و فرکانسی برای غنی‌تر کردن مجموعه ویژگی‌ها استفاده کنید.

۴. بصری‌سازی داده‌های صوتی:

- برای نمایش سیگنال‌های صوتی به صورت طیفی، از Spectrogram یا Mel-Spectrogram استفاده کنید.

- از روش‌های کاهش ابعاد مانند t-SNE یا PCA برای نمایش داده‌ها در فضای دوبعدی یا سه‌بعدی و تحلیل خوشه‌بندی یا طبقه‌بندی استفاده کنید.

۲.۶ طبقه‌بندی

همانطور که پیشتر ذکر شد، هدف ما در این قسمت، طبقه‌بندی استفاده از ویژگی‌های استخراجی مرحله قبل است. در مراحل مختلف طبقه‌بندی، مراحل زیر را در نظر داشته باشید. در این قسمت دو تسک مختلف تشخیص جنسیت گوینده صدا و تشخیص هویت آن همانطور که در مرحله قبل آشنا شدید به صورت closed-set authentication باید انجام شود. ویژگی‌هایی که در هر تسک استفاده می‌شود می‌تواند متفاوت باشد.

- تشخیص جنسیت (gender classification)

در این مرحله لازم است از صدای حداقل ۲۵ درصد افراد موجود در دیتاست استفاده کنید ولی می‌توانید تنها از بخش یا بخش‌هایی از صدای آن‌ها که در مراحل قبل جدا کرده‌اید استفاده کنید. همچنین در این بخش توجه کنید که نسبت دو کلاس به میزان یکسان باشد.

- تشخیص هویت (closed-set authentication)

در این روش شما باید داده‌های مربوط به صدای حداقل ۶ نفر (بصورت متوازن بین کلاس‌ها) را به صورت تصادفی انتخاب کرده و سپس به آموزش مدل بر اساس آنها بپردازید. به طور کلی باید این روند ۳ مرتبه برای حداقل ۶ فرد تصادفی متفاوت انجام شود. در نهایت گزارش شما باید شامل نتایج و تحلیل هر سه دسته باشد.

مراحل انجام هر تسک به شرح زیر است:

۱. داده‌ها را به دو دسته آموزش و تست تقسیم کنید. داده تست باید حداقل 25 درصد کل داده باشد. برای جداسازی داده تست و آموزش دقت داشته باشید که نسبت دو کلاس در هر دو دسته از داده‌ها، به یک میزان باشد.

۲. در این مرحله باید با استفاده از روش‌های یادگیری ماشین که در طول ترم آموخته‌اید، طبقه‌بندی را انجام دهید. در این مرحله **حداقل از سه** روش جداگانه استفاده کرده و نتایج را با هم مقایسه و تحلیل کنید. در این بخش صرفاً کدها به تنهایی حائز اهمیت نیستند بلکه در کنار آن‌ها تحلیل نمودارهای داده

وهمچنین مقایسه‌ی روش‌های مختلف، اهمیت ویژه‌ای دارد. می‌توانید برای طبقه‌بندی از روش‌های Logistic Regression⁵، MLP⁴، KNN³، SVM³، AdaBoost، XGBoost یا هر روش دیگری استفاده کنید.

۳. در هر کدام از روش‌ها، می‌توانید از روش‌های Normalization نیز استفاده کنید.

۴. برای هر طبقه‌بندی که مورد استفاده قرار می‌دهید، ماتریس آشفتگی، ROC Curve، تحلیل میزان خطای

هر کلاس و سایر موارد مورد نیاز را در گزارش خود بیاورید.

۲.۷ خوشه‌بندی

در اینجا همانند قسمت قبلی باید با استفاده از روش‌هایی که در درس آموخته‌اید، با انتخاب حداقل دو روش، داده‌ها را خوشه‌بندی کنید. با استفاده از روش‌هایی مثل silhouette score تعداد خوشه‌های مناسب را پیدا کرده و نمودار آن را رسم کنید. سپس خوشه‌بندی را به ازای تعداد خوشه مناسب که بدست آوردید و همچنین به ازای ۲ مقدار دیگر برای تعداد خوشه انجام دهید و نتایج حاصل را تحلیل و بررسی کنید. همچنین شباهت داده‌ی درون یک خوشه و تفاوت بین خوشه‌ها و دلایلی که برخی داده‌ها در یک خوشه قرار گرفته‌اند باید بررسی دقیق شوند. تحلیل و گزارش در این بخش از اهمیت بالایی برخوردار است.

³ Support vector machines

⁴ K-nearest neighbors

⁵ Multilayer perceptron

فصل ۳

گزارش کار

همانطور که قبلا هم گفته شد، علاوه بر کد درست، گزارش کار مفصل و توضیح و تحلیل درست داده و نمودارها از اهمیت بسیار بالایی برخوردار است. سعی کنید تمام نکات قابل ذکر در انجام پروژه را در گزارش کار ذکر کنید. دقت داشته باشید که گزارش کار اولیه، که مهلت آن تا ۳۰ ام آذر است، به عنوان مقدمه گزارش کار نهایی است. گزارش نهایی هم همراه با کد و تحلیل موارد خواسته شده تا ۲۵ دی ماه فرصت دارد. در زیر نکاتی را متذکر می شویم که حتما باید در گزارش کار نهایی ذکر شود.

- توضیح کامل نحوه کارکرد روش هایی که برای قسمت پیش پردازش استفاده کردید.
- تحلیل های لازم برای بخش خوشه بندی با تعداد خوشه هایی متفاوت.
- توضیح درباره علت انتخاب مدل هایی که برای طبقه بندی استفاده کردید.
- ذکر معیار هایی نظیر precision، recall، F1 score و غیره، برای هر کدام از مدل هایی که آموزش داده اید.
- معیار نهایی نمره دهی گزارش کار بوده و گزارش کار باید شامل تمامی بخش های مهم کد و توضیحات آنها باشد و در صورت نبودن توضیحات کد در گزارش کار نصف نمره آن بخش را نمی گیرید.
- گزارش کار میبایست مرتب و بخش بندی شده باشد و تحلیل های دقیق داشته باشد.
- کدهای خود را به صورت عکس در داخل گزارش کپی نکنید و با فرمتی مناسب متنی در گزارش قرار دهید .
- گزارش اولیه و نهایی میتواند به زبان فارسی و یا انگلیسی باشد.

فصل ۴

بارم بندی

از ۱۰۰	نمره دهی
۲۰	گزارش اولیه
۳۰	پیش پردازش و استخراج فیچر
۳۰	طبقه بندی
۲۰	خوشه بندی

فصل ۵

نکات پایانی

- هیچگونه شباهتی در انجام این پروژه بین افراد مختلف پذیرفته نمی شود. در صورت کشف هرگونه تقلب، مطابق قوانین درس با افراد خاطی برخورد خواهد شد.
- استفاده از مراجع با ارجاع به آنها بلامانع است. اما در صورتی که گزارش شما ترجمه عینی از آن ها باشد، یا از گزارش افراد دیگر استفاده کرده باشید کار شما تقلب محسوب می شود.
- بعد از مطالعه ی کامل و دقیق این توضیحات، در صورتی که سوالی در مورد پروژه داشتید به طراحان پروژه ایمیل بزنید.
- ایمیل دستیاران آموزش:

کیانا هوشانفر: k.hooshanfar@ut.ac.ir

رضا رضائیان: r.rezaeian@ut.ac.ir

محمد ولی نژاد: m.valinezhad@ut.ac.ir