

گزارش پروژه پایانی درس شاسائی الگو

جداسازی صدا بر اساس جنسیت و تحلیل خوشه بندی داده ها

مطهره پوررحیمی و نیوشا میر حکیمی

۸۱۰۱۹۶۴۳۴ و ۸۱۰۱۹۶۵۶۹

## مرتب سازی و تمیز سازی داده ها

در این مرحله دو فولدر ۱ و ۶ را که هر یک ، یک شماره سبجکت میانی نداشتند ( تعداد سبجکت های گرفته شده یکی کمتر از شماره آخرین سبجکت بود. ) ترتیب شان را درست کردیم. فایل های csv. را نیز یک مرتب سازی کردیم و به شکل سه ستون Subject که شماره سبجکت ها در آن قرار داشت، Gender که M نماینده مرد و F نماینده زن قرار دادیم تا برای خواندن فایل ها و لیبل زدن راحت تر باشد و Birth که سال تولد سبجکت بود.

## استخراج ویژگی

کد این بخش با نام load\_code.py در فولدر پروژه ذخیره شده است. در این کد استخراج ویژگی ها انجام شده است. برای استخراج ویژگی با توجه به منابع ، از ۵ گروه ویژگی زیر که هر یک به اختصار توضیح داده شده اند، استفاده کردیم. برای استخراج آن ها از کتابخانه librosa پایتون استفاده کردیم.

۱. MFCC (Mel-Frequency Cepstral Coefficients): ضرایبی هستند که می توانند Mel-Frequency Cepstrum (MFC) را تشکیل دهند. در واقع یک طیف غیر خطی از یک طیف دیگر هست که طیف اولیه MFC می باشد که یک بازنمایی از short term power spectrum یک سیگنال صوت است. ۴۰ فیچر به دست می دهد.
۲. Chromagram: ارتباط نزدیکی با ۱۲ تا کلاس گام صدا (pitch) دارد و ۱۲ پروفایل گام صدا را از هم متمایز می کند. ۱۲ فیچر به دست می دهد.
۳. Mel-Spectrogram: همان طیف توان است با این تفاوت که محور فرکانسی آن به صورت یک اسکیل غیر خطی شده، mel است. ۱۲۸ فیچر به دست می دهد.
۴. Contrast: پیک طیف، دره طیف و اختلاف این دو در هر زیر باند فرکانسی را بیان می کند. ۷ فیچر به دست می دهد.
۵. Tonnetz: ۶ فیچر به دست می دهد.

برای یک دست سازی فایل ها برای هر فولدر، تمام سبجکت ها را در فولدر Voices قرار دادیم. سپس آدرس هر داده ویس را در یک لیست ذخیره کرده و هم زمان شماره سبجکت و لیبل و سال تولد آن نیز در لیست های دیگر ذخیره سازی شدند. در نهایت دیتا فریمی ساخته شده از آدرس هر ویس، لیبل، سال تولد، شماره هر فولدر، و شماره سبجکت در هر فولدر ذخیره شد. سپس دیتافریم به تابع استخراج ویژگیها داده شد و از ستون آدرس برای لود دیتا استفاده شد. به دلیل ایراد فرم ذخیره سازی فایل های ۸ و ۹ و ۱ از این سه فایل داده استفاده نشده است.

پس از آماده سازی ماتریس داده ها به صورتی که در بالا ذکر شد، نمونه ها را شافل کرده و ۷۰٪ نمونه ها را مجموعه داده آموزش، ۲۰٪ را مجموعه داده validation و ۱۰٪ را نیز برای تست در نظر گرفتیم. داده های تست را برای evaluation نهایی کنار گذاشتیم و با استفاده از داده های آموزش طبقه بند های مورد نظر را، آموزش داده و برای tune کردن hyper parameter ها از مجموعه داده validation استفاده کردیم.

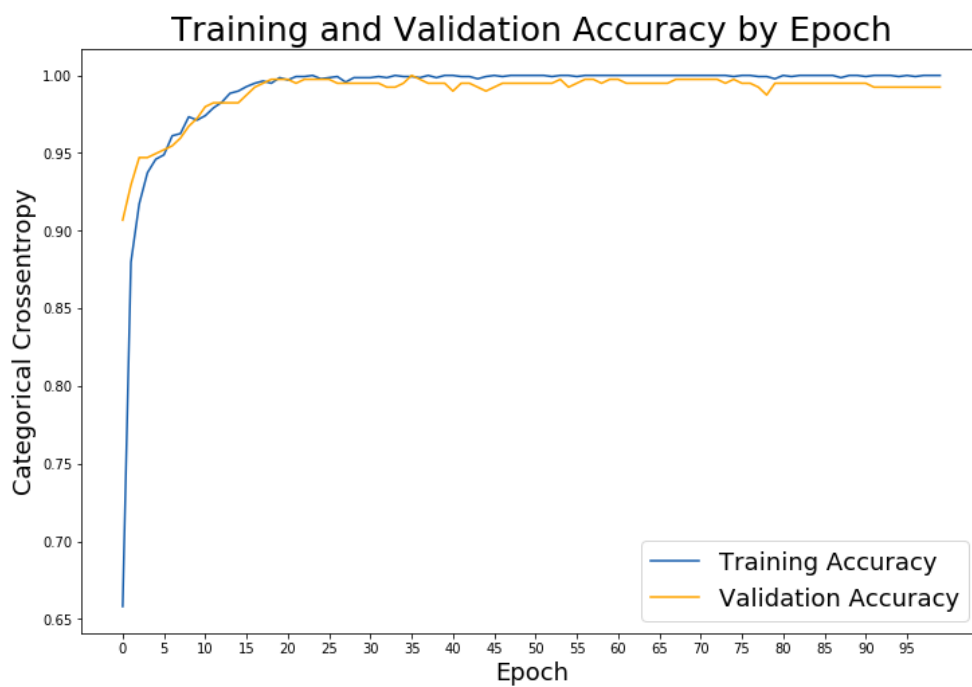
## طراحی طبقه بند

سه نوع طبقه بند استفاده کردیم :

۱. Neural network
۲. SVM (Linear, polynomial)
۳. Naïve Bayes

## ۱. شبکه عصبی

با استفاده از فیچرهای استخراج شده یک شبکه عصبی را آموزش داده و به عنوان طبقه بند از آن استفاده کردیم. در شکل ۱ accuracy دیتای آموزش و validation در طول epoch ها نمایش داده شده اند.



شکل ۱

Accuracy برای دیتای تست: 0.9898

که دقت بالاتری از تخمین قبلی ماست و نسبتاً دقت خوبی است.

## ۲. SVM

الف. Linear SVM:

نتایج روی دیتای validation: (البته tune کردن پارامتر نداشت و میتوانستیم دیتای تست را با validation مخلوط کنیم. اما برای حفظ consistency این کار را نکردیم.)

ماتریس confusion:

162	16
10	172

Accuracy = 0.927

نتایج روی دیتای تست:

ماتریس confusion:

167	5
11	182

Accuracy = 0.956

بنابراین با SVM خطی نیز به دقت خوبی رسیدیم.

ب. SVM with polynomial kernel

نتایج روی دیتای validation:

ماتریس confusion:

174	4
4	178

Accuracy = 0.977

نتایج روی دیتای تست:

ماتریس confusion:

169	3
6	187

Accuracy = 0.975

بنابراین با SVM با کرنل چند جمله ای نیز به دقت خوبی رسیدیم. هر چند چون با خطی به دقت خوبی رسیده بودیم خیلی نیازی به تست کردن با کرنل نبود.

۳. Naïve Bayes

نتایج روی دیتای validation:

ماتریس confusion:

60	118
11	171

Accuracy = 0.6416

نتایج روی دیتای تست:

ماتریس confusion:

57	115
13	180

Accuracy = 0.649

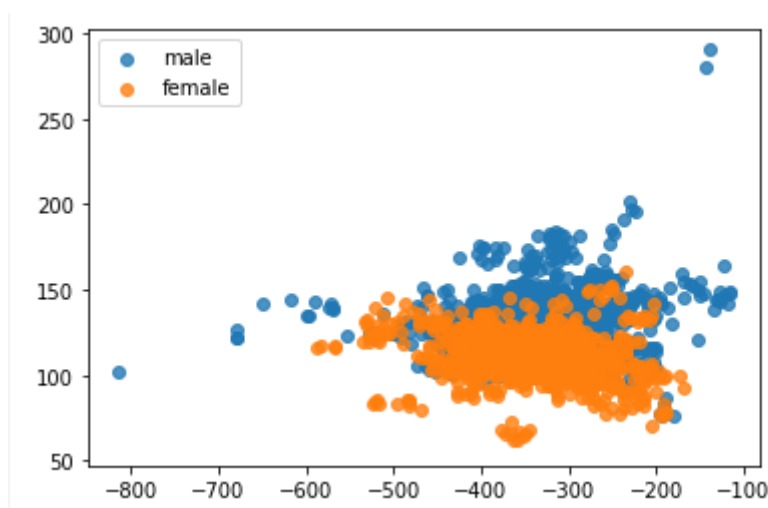
دقت کمتر از بقیه طبقه بند هاست.

## ارزیابی سه مدل و مقایسه آن ها

طبق نتایجی که هر یک از کلسیفایر ها روی مجموعه داده تست داشتند، شبکه عصبی بهترین عملکرد را داشته، بعد SVM با اختلاف کم و در نهایت Naïve Bayes با دقت پایین در رده آخر قرار گرفت.

## PCA

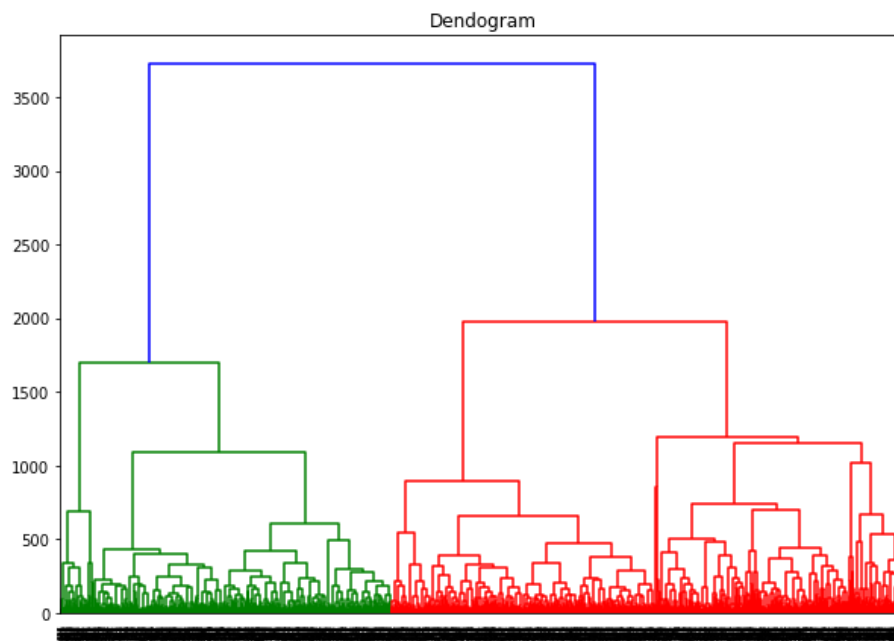
برای تبدیل فضای ۱۹۳ تایی فیچر ها به فضای کوچک تر برای این که بتوانیم باز نمایی بهتری از داده ها داشته باشیم (شکل ۲) ، از PCA استفاده کردیم. البته باید گفت به دلیل اینکه زمان ران شدن الگوریتم های کلسیفایر ها با همین ۱۹۳ فیچر کم بوده و به دقت های مورد نظر رسیدیم از این PC ها برای طبقه بندی مجدد استفاده نکردیم.



شکل ۲. بازنمایی داده ها در دو بعد با استفاده از تحلیل مولفه های اصلی

## خوشه بندی

برای خوشه بندی داده ها ابتدا دندوگرام رسم شد که در شکل ۳ مشاهده می شود همان طور که در شکل میبینیم ( قرمز و سبز ) که تعداد بهینه خوشه ها بر اساس تعداد خوشه هایی که بیشترین لایف تایم را دارند، ۲ است. و بر اساس لایف تایم آن ها تعداد کلاستر های بهینه برابر ۲ انتخاب می شود و سپس از دو الگوریتم سلسله مراتبی: Agglomerative و مبتنی بر بهینه سازی (Optimization): K-means استفاده کردیم برای خوشه بندی به تعداد دو خوشه استفاده کردیم.



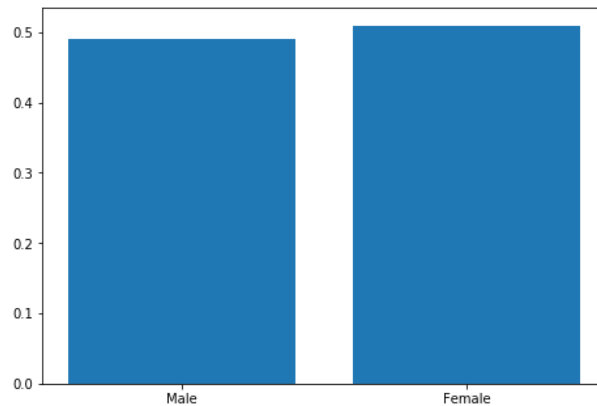
شکل ۳. دندوگرام

برای بررسی اینکه آیا این دو خوشه بیانگر دو جنسیت صدا هستند یا خیر، لیبل‌های به دست آمده را به لیبل‌های جنسیت داده‌ها مقایسه کردیم. ۰ نماینده مرد و ۱ نماینده زن بود. با این معیار و با پیاده‌سازی **Agglomerative** شباهت حدود ۶۱٪ بین خوشه بندی و ویژگی جنسیت مشاهده شد. با این معیار و با پیاده‌سازی **K-means** شباهت حدود ۵۲٪ بین خوشه بندی و ویژگی جنسیت مشاهده شد.

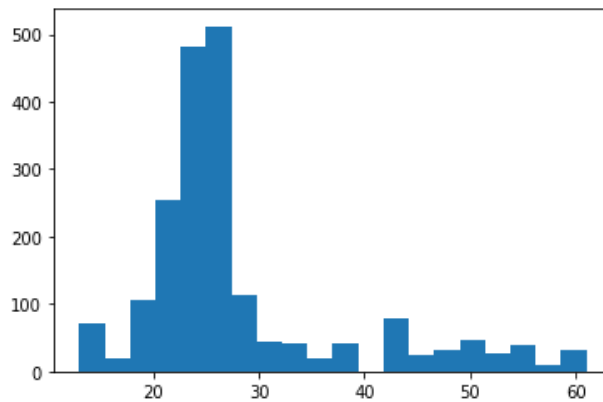
برای اینکه بررسی کنیم خوشه‌های کوچکتر تشکیل شده (تعداد بالاتر خوشه‌ها) نمایانگر چه چیز هستند، با استفاده از الگوریتم **Agglomerative** خوشه بندی با تعداد ۱۹۵ (تعداد کل افرادی که صدا هایشان را داشتیم). انجام دادیم. سپس لیبل‌های حاصل را با لیبل‌های **subject** ها بررسی کردیم تا ببینیم آیا نمایانگر اشخاص مختلف هستند یا خیر. به منظور این بررسی، لیبل‌های حاصل از کلاسترینگ را ستون اول و ستون دوم را شماره سابجکت‌ها گذاشتیم (ابتدا داده‌ها و شماره سابجکت‌ها را هم زمان و به یک ترتیب شافل کرده بودیم). سپس این ماتریس را با توجه به مقادیر ستون اول سورت کردیم. به این ترتیب کافی است تا ببینیم آیا به ازای یک گروه زیر هم در ستون اول که یکسان‌اند، اعداد مقابل آن‌ها در ستون دوم نیز مشابه هستند یا خیر. برای این کار لیستی ایجاد کردیم از ۱۱۵ لیست به ازای هر یک از کلاسترها. سپس سائز هر کلاستر را در بردار **sizes\_of\_clusters** ذخیره کرده و در هر کلاستر، سابجکت با بیشترین تکرار را به عنوان سابجکت غالب در **num\_of\_frequent\_subject** ذخیره کردیم. هم چنین تعداد تکرار این سابجکت در آن کلاستر را هم در **most\_frequencies\_in\_clusters** ذخیره کردیم تا با به دست آوردن نسبتش به تعداد اعضای آن کلاستر معیاری از هم‌گونی آن کلاستر را به دست آوریم، آن را **cluster signficancy** نام گذاشتیم. تعداد کلاسترهایی که این شاخص برای آن‌ها بیش از ۵۰ درصد است را برابر ۵۰ به دست آوردیم. یعنی از این ۱۱۵ تا خوشه، ۵۰ تا را میتوان به عنوان نماینده اشخاص متفاوت دید. پس روند به سمت جداسازی افراد می‌رود، اما دلالی از قبیل کافی نبودن داده برای هر نفر و تعداد متفاوت داده‌های هر کس (بعضی فقط ۳ تا) باعث این کمی دقت میشود. البته که استخراج فیچرهای متفاوت ممکن است این خوشه بندی را دقیق‌تر کند. (از لحاظ جدا سازی افراد)

## تحلیل آماری توزیع داده ها از نظر سن و جنسیت

در شکل ۴ نمودار میله ای نمایانگر دو جنسیت زن و مرد مشاهده میشود که همگونی خوبی را در دیتا به نمایش می گذارد. به دلیل مشکل در خواندن لیبل های جنسیت در فولدر های ۹۸ و ۹۱، همان طور که در بخش تمیز سازی ذکر شد که داده های این فولدر ها را وارد تحلیل ها نکردیم، اینجا نیز در تحلیل توزیع جنسیتی داده ها نیز لحاظ نشدند. اما در شکل ۵ که هیستوگرام توزیع سنی داده هاست، کلا داده ها لحاظ شدند.



شکل ۴. توزیع جنسیتی داده ها



شکل ۵. توزیع سنی داده ها

## منابع :

1. [https://www.projectrhea.org/rhea/index.php/Male\\_vs.\\_Female\\_Voice\\_characteristics](https://www.projectrhea.org/rhea/index.php/Male_vs._Female_Voice_characteristics)
2. <https://www.sciencedirect.com/science/article/abs/pii/S0892199796800191>
3. <https://www.hindawi.com/journals/sp/2019/7213717/>
4. <https://towardsdatascience.com/voice-classification-with-neural-networksff90f94358ec>
5. <http://www.primaryobjects.com/2016/06/22/identifying-the-gender-of-a-voice-using-machine-learning/>