به نام خدا

دانشکده علوم و فنون نوین

پردازش گفتار (۱۴۸-۰۵-۸۳)

نیم‌سال دوم ۱۴۰۲-۱۴۰۱

تمرین شماره ۱

تاریخ تحویل: ۱۴۰۱/۱۲/۲۲

---

۱. (۱۰٪) [پژوهش]

همان‌گونه که در کلاس بیان شد، پردازش گفتار در کاربردهای مختلف کاربرد دارد. هدف این سوال مروری بر برخی از کاربردهای دیگر پردازش گفتار است که در اسلایدهای کلاس ارائه نشده است. هر دانشجو دست‌کم یک موضوع (کاربرد یا روش) را بررسی کند. پس از مطالعه منابع مرتبط (مقاله، پایان‌نامه، گزارش و ...) از آنها یک گزارش کوتاه تهیه کنید. در این گزارش کوتاه باید به چیستی موضوع، روش و مدل بکار بسته شده، دادگان، نتایج اشاره کند (و هرچیز پایه‌ای به فراخور بودن در منبع آن موضوع برگزیده شده). توجه شود که برای [هر] موضوع، دست‌کم یک منبع معتبر را که در گزارش خود از آن بهره برده‌اید، همراه پاسخ بفرستید.

در انجام این بخش کوشا باشید و دقت به خرج دهید. چه بسا در برگزیدن موضوع پروژه نهایی‌تان کمک‌کننده باشد.

---

۲. (۱۵٪) [آشنایی با یک سیستم بازشناسی گفتار]

در این سوال می‌خواهیم با یک سیستم بازشناسی گفتار آشنا شویم. برای این کار با انتخاب‌های که در تلفن‌های هوشمند خود یا... دارید می‌توانید این بخش را انجام دهید. پیشنهاد ما به Google Speech Service است. اگر تلفن‌های هوشمند اپل دارید، این بخش را در یک تلفن هوشمندی که اندرویدی است انجام داده و گزارش دهید.

الف- تلاش کنید تا در شرایطی تقریبا ایده‌آل (نگفتن واژگان سنگین و غریب، نبود نویزهای محلی و...) ۵ جمله با طول حدودی ۱۰ واژه را به سامانه برگزیده شده بگویید تا آن‌ها را برای شما تایپ کند. اصل جمله‌های خود را با آنچه سیستم به شما بر می‌گرداند، مقایسه کنید و در گزارش خود بیاورید. همچنین در مورد چرایی رخداد WER در یک سیستم بازشناسی گفتار، پژوهشی انجام دهید.

ب- خطاهای رخ داده‌شده را برای هر واژه در جمله‌ای که گفته‌اید را با سنجه Word Error Rate (WER) گزارش کنید.

h.veisi@ut.ac.ir

یادداشت: *Word Error Rate، نرخ خطاهای رخ داده‌شده در رونویسی (transcription) بدست آمده از یک سیستم بازشناسی گفتار است. این خطاها می‌تواند نادرست نویسی واجها در یک واژه (S)، درج نادرست واژه‌های نگفته شده (I) و واژگان حذف شده (D) باشد. بنابراین:*

$$WER = \frac{D + I + S}{X}$$

در این رابطه، X شمار واژگان گفته‌شده است.
ازآنجایی که با یک نرخ سر و کار داریم، در صورت رخدادن خطا، برای هر واژه نهایتا عدد یک را در نظر بگیرید.

پ- در تلاشی دیگر، یک جمله بگویید و بعد آن را اصلاح کنید. رفتار سیستم را برای این حالت گزارش دهید. این کار را با یک چت‌بات (اگر با ویندوز هستید، Cortana و اگر با مک هستید، Siri) نیز انجام دهید (تنها این بخش چت‌بات می‌تواند به زبان انگلیسی باشد) و نتیجه را گزارش دهید.

۳. (۱۰٪+۶۵٪) [پیاده‌سازی: دسته‌بندی صداهای محیطی]

در این سوال می‌خواهیم یک برنامه بنویسید که بتواند صداها را با استخراج برخی ویژگی‌ها دسته‌بندی بکند. دادگان برای انجام این بخش از لینک زیر قابل دریافت است.

https://www.kaggle.com/datasets/mmoreaux/environmental-sound-classification-50

در این دادگان، ۵۰ دسته فایل صدا با نرخ نمونه‌برداری 16KHz قرار دارد. هر دسته دارای ۴۰ فایل صدا به طول ۵ ثانیه است. مسئله دسته‌بندی صداها بیشتر با ویژگی‌هایی از سیگنال گفتار بنام MFCC انجام می‌گیرد. در اینجا بجای MFCC، از سه ویژگی **Energy**، **Zero Crossing Rate** و **Spectral centroid** بهره ببریم. برای بدست آوردن این ویژگی‌ها، کتابخانه librosa می‌تواند کمک‌کننده باشد (در مورد این سه ویژگی، سر کلاس TA، صحبت خواهد شد). حال بنا داریم تا با بهره‌گیری از الگوریتم‌هایی که در درس یادگیری ماشین داشتیم کار دسته‌بندی را انجام دهیم.

الف) (۱۰٪+۳۵٪) مدل Logistic Regression

در این بخش، می‌خواهیم تا با بکارگیری مدل Logistic Regression، کار دسته‌بندی را انجام دهید. برای

h.veisi@ut.ac.ir

این بخش باید:

گام ۱- دادگان خوانده شود.

گام ۲- تابعی برای استخراج این سه ویژگی ها فایل‌های صدا بسازید و دادگان را فراهم کنید.

گام ۳- به صورت تصادفی، ۸۰٪ از دادگان برای آموزش و ۲۰٪ باقیمانده برای آزمون مدل تقسیم شود. دقت کنید که مدل در آموزش خود نباید از این ۲۰٪ داده را ببینند.

گام ۴- مدل Logistic Regression را بدون بهره‌گیری از کتابخانه‌هایی که این مدل را دارند، با رویکرد گرادیان کاهشی، آموزش دهید و ماتریس درهم‌ریختگی و از آن، معیارهای F1 score، Recall و Precession را گزارش دهید. در این آموزش، تعداد iteration را ۱۰۰۰ و نرخ یادگیری را ۰.۰۱ بدانید.

امتیازی (+۱۰٪): برای کار دسته‌بندی پیشنهاد شده، بجای بهره‌گیری از سه ویژگی یادشده، از MFCC بهره ببرید. MFCC و MFCC_mean را می‌توانید با قطعه کد زیر بدست آورید.

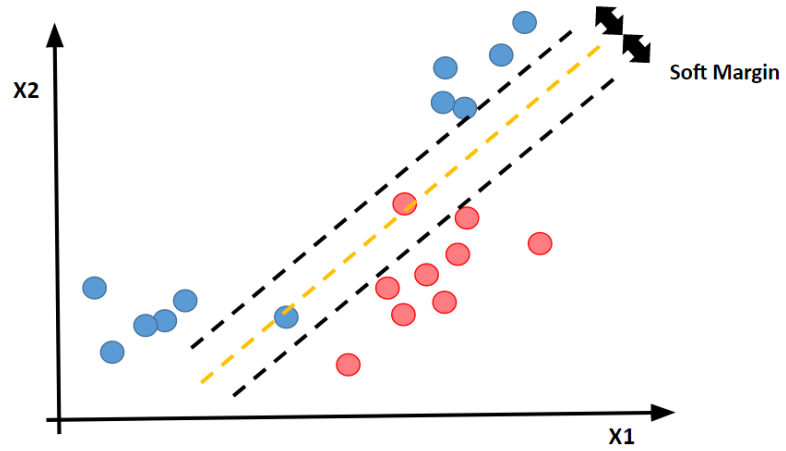
```
n_mfcc = 13
MFCC = librosa.feature.mfcc(signal, n_fft= number_of_samples_per_fft,
                            hop_length=shift_value, n_mfcc= n_mfcc)
MFCC_mean = np.maen(mfcc, axis=1)
```

ب) (۳۰٪) مدل SVC

ماشین بردار پشتیبانی (SVM) یکی از روش‌های یادگیری با نظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. هدف این الگوریتم پیدا کردن ابرصفحه‌ای برای جدا کردن کلاس‌ها از یکدیگر می‌باشد که از دو رویکرد حاشیه سخت (Hard-Margin) و حاشیه نرم (Soft-Margin) استفاده می‌کند.

h.veisi@ut.ac.ir

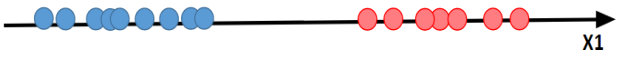
استفاده از روش حاشیه سخت می‌تواند در مواردی محدودیت‌هایی ایجاد کند که یکی از آنها وابستگی شدید جداکننده به داده‌های مرزی است که با وجود داده‌های نویزی به دلیل وابستگی بسیار زیاد به داده‌های مرزی منجر به overfit شدن مدل می‌شود به همین دلیل از رویکرد حاشیه نرم استفاده می‌شود که درشکل زیر می‌توان مشاهده کرد.



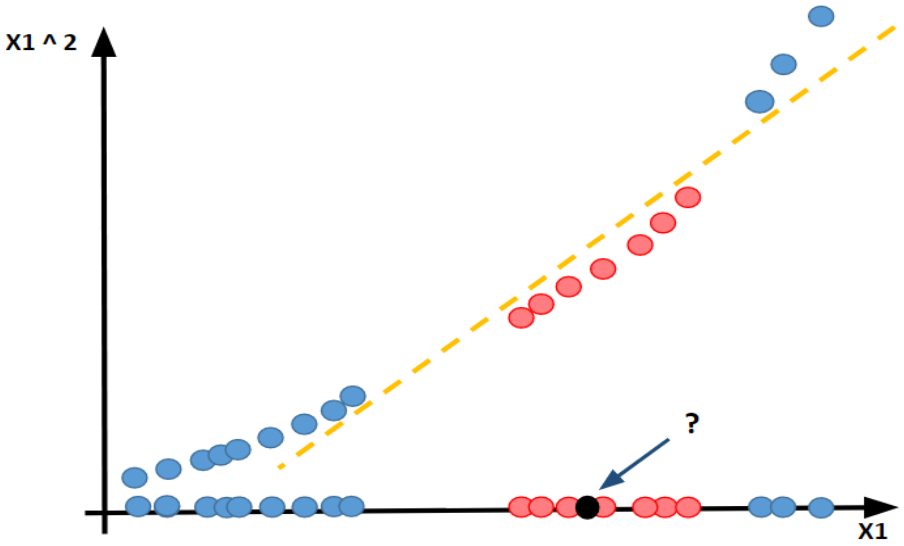
شکل ۱- soft margin

این الگوریتم همواره با داده‌ها در ابعاد پایین شروع می‌کند و داده‌ها را به ابعاد بالاتر منتقل می‌کند که بتواند راحت تر ابر صفحه مورد نظر را پیدا کند برای ارتقاء ابعاد داده می‌توان از توابع مختلف مانند چندجمله ای درجه n استفاده کرد (Polynomial Kernel) که برای مثال می‌توان در شکل زیر مشاهده کرد. که داده‌ها با ابعاد پایین به ابعاد بالاتر منتقل شده است و می‌توان ابرصفحه خواسته شده را راحت‌تر پیدا کرد.

h.veisi@ut.ac.ir



شکل ۲ - داده‌ها با ابعاد پایین



شکل ۳- انتقال داده‌ها به ابعاد بالاتر

پس یکی از مهم‌ترین ابرپارامترهای مهم تعیین نوع هسته و تعیین ابرپارامترها برای هسته مورد نظر می‌باشد برای مثال، می‌توان حل مسئله طبقه‌بندی را با درجه‌های مختلف در ابعاد بالا در شکل زیر مشاهده کرد.

h.veisi@ut.ac.ir

به نام خدا

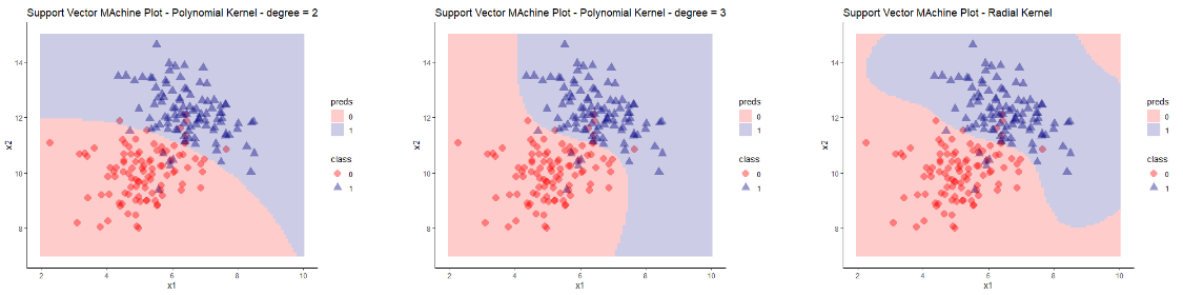
دانشکده علوم و فنون نوین

پردازش گفتار (۱۴۸-۰۵-۸۳)

نیم‌سال دوم ۱۴۰۲-۱۴۰۱

تمرین شماره ۱

تاریخ تحویل: ۱۴۰۱/۱۲/۲۲



شکل ۱ حل مسئله طبقه‌بندی با ابعاد مختلف

ابرپارامتر بعدی گاما است که تفاوت گاما بزرگ و کوچک را می‌توان در شکل زیر مشاهده کرد.

شکل ۲ تفاوت پارامتر گاما بزرگ و گاما کوچک در یادگیری

در این مسئله خواسته شده است که دانشجویان همانند قسمت قبل از ویژگی خواسته شده استفاده کنند و با الگوریتم SVM برای طبقه‌بندی یا SVC صوت‌های مختلف را دسته‌بندی کنند و برای هسته از Polynomial استفاده شود.

در قسمت اول از این مسئله از دانشجویان خواسته می‌شود با تعیین درجه مناسب و مقدار گاما مناسب الگوریتم یادگیری انجام شود و در قسمت دوم ارزیابی برای داده‌های تست انجام شود و برای ارزیابی از recall،precision،accuracy، f1 score گزارش شود و همچنین نمودار برای ۳ حالت مختلف از درجه‌های مختلف polynomial به انتخاب دانشجو نسبت به accuracy رسم شود و همچنین در انتها ماتریس در هم‌ریختگی گزارش شود (می‌توان برای پیاده‌سازی از کتابخانه‌های آماده استفاده شود)

h.veisi@ut.ac.ir