سوالات بخش عملی

پیش پردازش و استخراج ویژگی

۱. به نظر شما قطعه‌بندی داده‌ها برای این دیتاست مفید است؟ چرا؟

داده‌ها در این پروژه تعدادی فایل صوتی هستند که در هرکدام فقط یک رقم توسط یک گوینده خوانده می‌شود. در واقع هر فایل صوتی فقط شامل یک بخش است که در آن بخش دو فاکتور رقم و گوینده وجود دارد. در نتیجه قطعه‌بندی داده‌ها به صورتی که هر فایل صوتی به بخش‌های جداگانه تقسیم شود در بخش تشخیص رقم ممکن نیست چون هر فایل صوتی فقط یک رقم دارد و قطعه کردن آن باعث از دست رفتن و تخریب داده‌ها می‌شود. در بخش تشخیص گوینده نیز به این علت که فایل‌های صوتی به اندازه کافی کوتاه هستند و تعداد آن‌ها هم مناسب train کردن یک مدل هست، قطعه‌بندی تاثیر مثبتی نخواهد داشت.

۲. در مورد هر کدام از این ویژگی‌ها تحقیق کنید و روابط بین آنها را توضیح دهید.

- MFCC: Mel Frequency Cepstral Coefficients خواصی از سیگنال‌های صوتی هستند که از تبدیل فوریه سیگنال به دست آمده و برای تشخیص گفتار و گوینده صوت، آنالیز موسیقی و تشخیص احساسات استفاده می‌شوند. MFCC به این دلیل مورد استفاده زیاد قرار می‌گیرد که روی ویژگی‌های مهم سیگنال صوتی تمرکز کرده و ویژگی‌های کمتر مهم آن را نادیده می‌گیرد.
- Zero Crossing Rate: نشان دهنده این است که سیگنال صوتی با چه فرکانسی علامت خود را تغییر می‌دهد و برای تشخیص تن صدا یا تشخیص ژانر موسیقی مورد استفاده قرار می‌گیرد.
- Mel-Spectrogram: فرکانس سیگنال صوتی را در طول زمان اندازه گیری می‌کند و فهم انسان از تن صدا را شبیه سازی می‌کند. این ویژگی هم برای تشخیص گفتار و تشخیص ژانر موسیقی استفاده می‌شود.
- Chroma Features: نشان دهنده میزان انرژی تن‌های صدای مختلف هستند و بلندی یا آرامی صدا در آن‌ها تاثیری ندارد. این ویژگی نیز برای تشخیص آکوردهای موسیقی، تشخیص ژانر موسیقی و تشخیص ملودی مورد استفاده قرار می‌گیرد.

MFCC و Mel-Spectrogram هر دو اطلاعاتی از طیف سیگنال صوتی به دست می‌دهند ولی ضرایب MFCC علاوه بر آن اطلاعاتی درمورد تجزیه طیفی سیگنال نیز می‌دهد. همچنین Zero Crossing Rate اطلاعاتی درمورد ویژگی‌های زمانی سیگنال می‌دهد که در کنار دو ویژگی دیگر می‌تواند برای تحلیل جنبه‌های پیچیده‌تری از سیگنال صوتی مفید باشد. همچنین Chroma Features اطلاعاتی را که با ویژگی‌های طیفی ثبت شده‌اند تکمیل می‌کنند. این ویژگی‌ها در کنار هم بهترین تحلیل از سیگنال صوتی را می‌دهند.

۳. robustness و حساسیت ویژگی‌های MFCCs را نسبت به تغییرات در سیگنال‌های صوتی بررسی کنید.

ویژگی‌های MFCC معمولا به عنوان ویژگی‌های ساختاری و مفید شناخته می‌شوند ولی نسبت به تغییرات در سیگنال‌های صوتی به ویژه تغییرات در شدت صوت و نویزهای محیط کمی حساس هستند.

- حساسیت نسبت به شدت صوت: به این علت که MFCC بر پایه لگاریتم اندازه‌گیری‌های طیفی محاسبه می‌شود، تغییرات خطی در شدت صوت می‌تواند تاثیر قابل توجهی در مقادیر MFCC ایجاد کند.
- حساسیت نسبت به نویزهای محیط: نویزهای محیط (مخصوصا نویزهای غیرخطی) می‌تواند در محتوای طیفی سیگنال تغییراتی ایجاد کند که باعث افت کارایی ویژگی‌های MFCC می‌شود.

برای کم کردن تاثیر این مشکلات می‌توان از راه حل‌هایی مانند فیلترینگ در پیش پردازش و یا نرمال‌سازی داده‌ها استفاده کرد.

۴. آیا موارد خاصی وجود دارند که ضرایب MFCC کارایی کمتری داشته باشند؟

در برخی موارد ضرایب MFCC عملکرد ضعیف‌تری دارند:

- سیگنال‌های با تنوع زیاد طیفی: در صورتی که داده‌ها شامل سبک‌های موسیقی مختلف یا سیگنال‌های صوتی از منابع مختلف باشد و شامل تنوع طیفی زیادی باشد ممکن است ضرایب MFCC کارایی کمتری داشته باشند.
- سیگنال‌های با نسبت سیگنال به نویز پایین: در سیگنال‌هایی که این نسبت (SNR) در آن‌ها پایین است نویزهای موجود ممکن است تغییر شدت طیف سیگنال را افزایش دهند و به علت حساسیت ضرایب MFCC به نویز این ضرایب در این موارد کارایی کمتری داشته باشند.
- سیگنال‌های با ویژگی‌های زمانی مهم: به این علت که ضرایب MFCC خواص زمانی سیگنال را نادیده گرفته و فقط اطلاعات طیفی را در نظر می‌گیرند، در سیگنال‌هایی که زمان در آن‌ها مهم است این ضرایب کارایی بالایی ندارند.

۵. چرا در محاسبه MFCC فریم‌های استفاده شده با یکدیگر هم‌پوشانی دارند؟

- حفظ اطلاعات زمانی: این کار باعث می‌شود مواردی مانند تغییرات ملودی، نویزهای کوتاه مدت و تغییرات دینامیکی در سیگنال به درستی تشخیص داده شود.
- تشخیص پیک‌های صوتی و تغییرات سریع: این امر امکان تشخیص مواردی مانند پیک‌های صوتی کوتاه مدت یا تغییرات سریع در فرکانس طیفی سیگنال را بیشتر می‌کند.
- کاهش تأثیرات لبه‌ها: وقتی که فریم‌ها با یکدیگر هم‌پوشانی دارند، لبه‌های سیگنال در فریم‌های مختلف نرم‌تر و اغلب به شکل میله می‌شوند. این باعث می‌شود که اثرات لبه‌ها در ضرایب MFCC کمتر قابل تشخیص باشد.

۶. چرا در اکثر پروژه‌های مرتبط با صوت تنها از ۱۲ یا ۱۳ ضریب ابتدایی MFCC استفاده می‌شود؟

ضرایب ابتدایی MFCC که معمولا از ۰ تا ۱۲ یا ۱۳ هستند، اطلاعات مهمی از سیگنال را ذخیره می‌کنندو باقی ضرایب نسبت به ضرایب اولیه از اهمیت کمتری برخوردار هستند. همچنین این امر باعث کاهش پیچیدگی داده‌ها شده و باعث می‌شود مدل ساخته شده در زمان کمتر عملکرد بهتری داشته باشد.

آشنایی با HMM

۱. توضیح دهید منظور از State ها و Observation چیست؟ در این تمرین State ها کدامند و Observation چگونه بدست می‌آید؟

- State: در HMM استیت‌ها دنباله‌ای از حالت‌ها هستند که هر استیت یک وضعیت مخفی است که سیستم می‌تواند در آن قرار داشته باشد.
- Observation: در HMM هر استیت می‌تواند یک یا چند Observation داشته باشد که مخفی نیست و بعنوان خروجی قابل مشاهده است.

در این تمرین State ها مفاهیمی انتزاعی و نمایانگر تاریخچه‌ای از چیزی که سیستم دیده است می‌باشد. هر زیردنباله از استیت‌ها ویژگی‌هایی از سیگنال‌های صوتی دیده شده در خود ذخیره می‌کنند. همچنین Observation ها ویژگی‌های استخراج شده از سیگنال‌های صوتی مانند MFCC است.

۲. مدل‌های HMM را میتوان بر اساس میزان وابستگی میان State های پنهان دسته‌بندی کرد، مدلی که در این تمرین به پیاده‌سازی آن می‌پردازید یک مدل First-Order HMM است. دلیل نامگذاری آن و همچنین ویژگی های آن را بررسی کنید و تفاوت آن با مدل های دیگر در این دسته‌بندی را بیان کنید.

در مدل First Order HMM هر استیت تنها به استیت قبلی خود وابسته است و Observation ها به صورت مستقل از یکدیگر به یک استیت وابسته هستند. به همین دلیل این مدل اینگونه نامگذاری شده است. تفاوت این مدل با مدل Second Order HMM در این است که در این مدل استیت‌ها به دو استیت قبل از خود وابسته هستند و این باعث می‌شود که مدل‌های Second Order انعطاف پذیری بیشتری در نمایش الگوهای پیچیده‌تر داشته باشند. همچنین مدل‌های دیگر Higher Order HMM وجود دارند که در آن‌ها هر استیت به تعداد بیشتری استیت قبل از خود وابسته است و برای نمایش الگوهای پیچیده‌تر مناسب‌تر هستند.

۳. درباره HMM تحقیق کنید و توضیح دهید که این مدل برای بررسی و تحلیل چه پدیده‌هایی مناسب است؟ چرایی این موضوع را توضیح دهید.

این مدل برای بررسی و تحلیل سیستم‌هایی که دارای اطلاعات غیرقابل مشاهده هستند استفاده می‌شود. برای مثال در زمینه‌های تحلیل سیگنال، تشخیص پترن، تشخیص گفتار یا موسیقی کاربرد دارد. به این دلیل که این مدل قابلیت مدل‌سازی فرایندهای پنهان و قدرت پیش‌بینی دارد برای تحلیل این‌گونه پدیده‌ها مناسب است.

۴. مدل HMM نیز مانند هر مدل دیگری دارای مزایا و معایبی است که آن را ویژه می‌کند. مزایا و معایب این مدل را بررسی کرده و هر کدام را مختصرا توضیح دهید.

مزایا:

- مدل‌های HMM انعطاف پذیری بالایی دارند و می‌توانند برای مدل کردن طیف گسترده‌ای از سیستم‌ها استفاده شوند.
- قابلیت مدل‌سازی الگوهای پنهان را دارند.
- دقت بالایی دارند.
- معمولا حساسیت کمی دارند.

معایب:

- مدل‌های HMM نیازمند تعیین تعداد ثابتی استیت هستند.
- فرضیات ساده‌سازی زیادی را درمورد داده‌ها انجام می‌دهد.
- وابسته به شروع اولیه استیت‌هاست و انتخاب اشتباه می‌تواند منجر به عملکرد نامناسب مدل شود.

۵. انواع مختلفی از مدل های HMM وجود دارد، درباره آن‌ها تحقیق کنید و چند مورد را بطور مختصر بررسی کنید.

- **Gaussian HMM:** این مدل فرض می‌کند که توزیع احتمال استیت‌ها و مشاهده‌ها گوسی است. این مدل‌ها برای داده‌های پیوسته و پراکنده مانند سیگنال‌های صوتی و تصویری مناسب هستند.
- Classification-based HMM: در این مدل هر استیت نمایانگر یک دسته‌بندی مشخص از داده‌هاست. این نوع از مدل‌ها معمولا برای دسته‌بندی داده‌ها استفاده می‌شوند.
- Time-Invariant HMM: این نوع مدل فرض می‌کند استیت‌ها و مشاهدات در طول زمان ثابت می‌مانند. این مدل برای داده‌هایی که الگوهای زمانی ثابت دارند، مناسب است.

Implementing from Scratch

ممکن است نتایج شما در بخش اول و دوم فرق کند و مدل آماده نتایج متفاوت و دقت بالاتری نسبت به مدل طراحی شده توسط شما داشته باشد. این اختلاف ممکن است چه دلایلی داشته باشد؟ درباره عوامل تاثیرگذار بر روی این اختلاف دقت تحقیق کنید.

نتایج بخش اول دقت نزدیک ۹۰ درصد و نتایج بخش دوم دقت کمتر از ۸۰ درصد دارند. دلیل این اختلاف می‌تواند موارد زیر باشد:

- کتابخانه hmmlearn از روش‌های پیچیده‌تری برای مقداردهی اولیه پارامترها استفاده می‌کند.
- این کتابخانه برای ساخت مدل‌های hmm ساخته شده و به این علت ممکن است از الگوریتم‌های پیشرفته‌تر و بهینه‌تری مانند الگوریتم Viterbi برای آموزش مدل استفاده کند.

ارزیابی و تحلیل

۱. درباره هر کدام از معیارهای بالا تحقیق کنید و نحوه محاسبه هر یک را توضیح دهید.

- Accuracy: Number of correctly classified samples / Total number of samples
- Precision:
  Number of true positive samples / Number of samples classified as positive by the model
- Recall: Number of true positive samples / Total number of positive samples
- F1-Score: 2*(precision * recall) / (precision + recall)

۲. آیا محاسبه معیارهای ذکر شده برای این پروژه که multi-class است، چالشی دارد؟ اگر بله چه راه حلی برایش دارید؟

محاسبه معیارهای ذکر شده برای مدلی که multi-class است سخت‌تر از محاسبه برای یک مدل دودویی است. راه حل‌های مختلف عبارتند از:

- میانگین وزن‌دار: در این رویکرد سهم هر دسته در میانگین با وزن‌هایی که در داده وجود دارند در نظر گرفته می‌شود. این روش زمانی مفید است که دسته‌ها نامتوازن هستند، زیرا به معیارهای دسته‌های بزرگتر وزن بیشتری داده می‌شود.
- Confusion Matrix: استفاده از این ماتریس شهودی بهتر از میزان درستی عملکرد مدل داده و با استفاده از آن می‌توان معیارهای گفته شده را محاسبه کرد.
- استفاده از معیارهای ویژه: برای مسائل چند دسته‌ای معیارهای ارزیابی ویژه‌ای وجود دارد مانند امتیاز F1 چند دسته‌ای، کاپا کوهن و ضریب همبستگی متیوز که به طور خاص برای این نوع از مسائل طراحی شده‌اند و ارزیابی جامع‌تری از عملکرد مدل را ارائه می‌دهند.

۳. توضیح دهید که هر کدام از معیارها چگونه مدل را ارزیابی میکنند.

- Accuracy: این معیار نشان دهنده درصد نمونه‌های صحیحی است که مدل به درستی دسته‌بندی کرده است.
- Precision: نشان می‌دهد که چه نسبتی از نمونه‌هایی که مدل به عنوان یک دسته خاص تشخیص داده است، واقعا به آن دسته تعلق دارند.
- Recall: نشان می‌دهد که مدل چه نسبتی از تمام نمونه‌های مثبت را شناسایی کرده است.
- F1: امتیاز F1 یک معیار کلی برای ارزیابی عملکرد یک مدل است که ترکیبی از Accuracy و Recall است. این معیار به عنوان میانگین هندسی از Accuracy و Recall محاسبه می‌شود و می‌تواند بهترین حالت بین این دو را نشان دهد.

۴. تفاوت میان Recall و Precision را بیان کنید و توضیح دهید چرا هر کدام به تنهایی برای ارزیابی مدل کافی نیست؟ برای هر یک مثالی بیاورید که در آن، این معیار مقدار بالایی دارد اما مدل عملکرد خوبی ندارد.

Recall و Precision دو معیار مهم برای ارزیابی عملکرد مدل‌های دسته‌بندی هستند، اما هر یک از آن‌ها به تنهایی نمی‌تواند تمام جنبه‌های عملکرد مدل را به خوبی ارزیابی کند. تفاوت اصلی میان این دو معیار در تمرکز آن‌ها است.

مثال Recall بالا: یک مدل تشخیص بیماری قلبی دارای Recall بالایی است، اما Precision پایینی دارد. این به این معنی است که مدل بسیاری از بیماران را به درستی شناسایی کرده است (Recall بالا)، اما همچنان تعداد زیادی از افراد سالم را به اشتباه به عنوان بیمار تشخیص داده است (Precision پایین).

مثال Precision بالا: یک مدل تشخیص اسپم ایمیل دارای Precision بالایی است، اما Recall پایینی دارد. این به این معنی است که بسیاری از ایمیل‌های شامل اسپم توسط مدل به درستی شناسایی شده‌اند (Precision بالا)، اما مدل بسیاری از ایمیل‌های اسپم را از دست داده و به اشتباه به عنوان ایمیل‌های معمولی تشخیص داده است (Recall پایین).

۵. معیار F1 از چه نوع میانگین‌گیری استفاده میکند؟ تفاوت این نوع میانگین‌گیری با میانگین‌گیری عادی چیست و در اینجا چرا اهمیت دارد؟

این معیار از میانگین هندسی استفاده می‌کند که به شکل زیر محاسبه می‌شود:

$$G = \sqrt[n]{a_1 \times a_2 \times a_3 \cdots a_n}$$

این نوع میانگین‌گیری ارزش بیشتری به مقادیر پایین می‌دهد به طوری که اگر یکی از مقادیر ۰ باشد کل میانگین ۰ می‌شود. این باعث می‌شود که F1 مدل‌هایی را که تفاوت بزرگی بین Precision و Recall دارند، مجازات کند و ارزیابی متوازنی از عملکرد مدل ارائه دهد.

۶. مدل خود را بر اساس رقم گفته شده در فایل صوتی آماده کنید. Confusion Matrix رسم کنید و دو معیار Accuracy و Precision را محاسبه کنید. در آخر مقادیر به دست آمده را تحلیل کنید.

هر دو معیار Accuracy و Precision مقادیر مناسب و بالایی دارند و این معیارها در مدل train شده توسط کتابخانه hmmlearn نسبت به پیاده سازی from scratch مقادیر بیشتری دارند. همچنین دیده می‌شود که تشخیص رقم ۳ از بین بقیه رقم‌ها سخت‌ترین بوده است.

۷. مدل خود را بر اساس گوینده آماده کنید. Confusion Matrix رسم کنید و دو معیار Accuracy و Precision را محاسبه کنید. در آخر مقادیر به دست آمده را تحلیل کنید.

هر دو معیار Accuracy و Precision مقادیر مناسب و بالایی دارند و این معیارها در مدل train شده توسط کتابخانه hmmlearn نسبت به پیاده سازی from scratch مقادیر بیشتری دارند. همچنین دیده می‌شود که تشخیص صدای jackson برای مدل سخت‌تر از تشخیص صدای دیگر گوینده‌ها بوده است.

۸. تفاوت نتایج بخش‌های ۶ و ۷ را بررسی کنید و علل آن را مشخص کنید.

همانطور که از مقادیر Accuracy مشخص است، دقت مدل‌ها در تشخیص گوینده‌ها بیشتر بوده ولی به طور عکس مقادیر Precision برای تشخیص ارقام بیشتر است. این ممکن است به دلیل توزیع کلاس‌ها و تعادل بین تعداد نمونه‌های مثبت و منفی در داده‌ها باشد.