آموزش پروژه‌محور پردازش صوت و گفتار با پایتون

فصل دوم: ماشین لرنینگ روی صدا

درس 5

استخراج ویژگی‌های پیشرفته برای ماشین لرنینگ

طراح و مدرس: مصطفی کرمانی‌نیا

تحت نظارت علمی: دکتر مهدی سیفی‌پور

Made with GAMMA

چرا MFCC به تنهایی کافی نیست؟

صدای انسان فراتر از حنجره است!

در فصل قبل، با MFCC توانستیم شکل حنجره (اثر انگشت) را پیدا کنیم. اما آیا این برای هوش مصنوعی کافیست؟ خیر!

احساسات و بافت صدا

هوش مصنوعی برای تشخیص اینکه یک نفر عصبانی است، دروغ می‌گوید یا خوشحال است، نیاز دارد "بافت"، "خراشیدگی" و "روشنی" صدا را هم درک کند.

ماموریت امروز

می‌خواهیم ویژگی‌هایی استخراج کنیم که مثل ادویه‌های جادویی، خوراکِ بی‌نظیری برای مدل‌های Machine Learning بسازند.

Made with GAMMA

نقشه راه: آماده‌سازی دیتا برای ماشین

گام اول: ZCR
شمارش تقاطع‌های صفر — تشخیص نویز از صدای صاف

گام دوم: Spectral Centroid
یافتن مرکز ثقل صدا — درک روشنی و بمی صدا

گام سوم: مهندسی ویژگی‌ها
تبدیل مفاهیم ریاضی به اعداد قابل فهم

گام چهارم: پل ارتباطی با AI
ساخت دیتافریم Pandas برای تزریق به مدل

گام پنجم: کدتایم!
تست زنده روی میکروفون در پایتون

Made with GAMMA

نرخ عبور از صفر (Zero-Crossing Rate) چیست؟

تعریف ساده

سیگنال در هر ثانیه چند بار خط صفر را قطع می‌کند؟ (از مثبت به منفی و برعکس می‌رود).

درک شهودی

صداهای نویزی و خشن، نوسانات بسیار سریع و نامنظمی دارند و مدام خط صفر را قطع می‌کنند. اما صداهای صاف و هارمونیک، موج‌های آرام‌تری دارند.

کاربرد اصلی

بهترین ابزار برای تفکیک حروف بی‌صدا (مثل س، ش، ف) از حروف صدادار (مثل آ، او).

Made with GAMMA

نرخ عبور از صفر در عمل

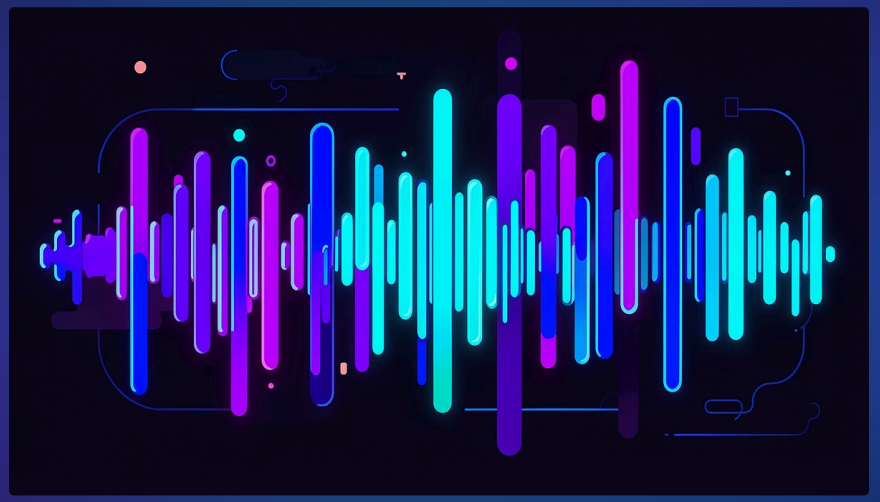
سناریوی کدتایم
در بخش عملی، به صورت لایو صدای خودمان را ضبط می‌کنیم.

تست اول: «اااااا»
خواهیم دید که نمودار ZCR در پایین‌ترین حد خود است.

تست دوم: «شـــششششش»
خواهیم دید که ZCR ناگهان منفجر می‌شود!

نکته طلایی
این دقیقاً همان منطقی است که هندزفری‌های هوشمند برای تشخیص شروع صحبت شما (VAD) استفاده می‌کنند!

Made with GAMMA



مرکز ثقل طیفی (Spectral Centroid) چیست؟

مرکز جرم صدا

همانطور که یک جسم فیزیکی نقطه تعادل (مرکز ثقل) دارد، طیف فرکانسیِ صدا هم یک "مرکز ثقل" دارد.

درک شهودی (Brightness)

این ویژگی مستقیماً "روشنی" یا "شفافیت" صدا را اندازه می‌گیرد. آیا صدا تیره و خفه است یا روشن و تیز؟

کلیات فرمول

میانگینِ وزن‌دار فرکانس‌ها؛ فرکانس‌هایی که انرژی (دامنه) بیشتری دارند، مرکز ثقل را به سمت خود می‌کشند.

Made with GAMMA

در عمل: سوت در برابر Centroid طبل!

صداهای بم

اگر با صدای کلفت و بم صحبت کنیم (مثل طبل)، مرکز ثقل روی فرکانس‌های پایین می‌ایستد.

صداهای زیر

اگر در میکروفون سوت بزنیم، مرکز ثقل به شدت بالا می‌رود!

کاربرد در AI

این ویژگی برای تشخیص "خشم" یا "هیجان" در پروژه‌ی دروغ‌سنج صوتی ما حیاتی است!

Made with GAMMA

پل ارتباطی: از صوت تا Pandas

مدل‌های هوش مصنوعی کور هستند!

الگوریتم‌هایی مثل Random Forest یا SVM فایل‌های .wav را نمی‌فهمند. آن‌ها فقط "جدول‌هایی از اعداد" (Tabular Data) را درک می‌کنند.

جادوی Pandas

ما باید تمام ویژگی‌هایی که یاد گرفتیم (MFCC, ZCR, Centroid) را در یک ساختار منظم (DataFrame) بسته‌بندی کنیم.

سوخت‌گیری کامل

هر ردیف از این جدول، نمایانگر یک فایل صوتی، و هر ستون نمایانگر یک ویژگی ریاضی است. حالا ماشین آماده یادگیری است!

Made with GAMMA

کدتایم!

سوخت‌گیری هوش مصنوعی در پایتون

استخراج ZCR و Centroid

کدنویسی ویژگی‌های پیشرفته با librosa و تست لایو میکروفون.

تست "ش" در برابر "آ"

رسم نمودارها برای درک تفاوت نویز و صدای هارمونیک.

ساخت Dataframe

تجمیع تمام ویژگی‌ها در Pandas برای استفاده در مدل‌های یادگیری ماشین آینده.

Made with GAMMA