برای دیتا ست ما GTZAN را انتخاب کرده ایم. خوبی این دیتا ست این است که دارای ده نوع ژانر است و حجم مناسبی دارد، اما بهای این حجم مناسب این است که برای هر ژانر تنها 30 ترک دارد و میدانیم که برای شبکه عصبی های عمیق این رقم مناسب نیست!

کاری که برای حل این مشکل میکنیم این است که هر ترک را به چندین قسمت تقسیم میکنیم تا داده ها را افزایش دهیم.

برای تبدیل داده ها به فرمت مناسب از mfcc استفاده میکنیم که قبلا مفصل راجبش حرف زدیم.

ابتدا یک دیکشنری می‌سازیم که دارای سه “key” است و “value” هر کدام یک لیست است:

“mapping”: دارای اسامی تمام ژانر های موجود در دیتا ست است.

“mfcc” : شامل تمام mfcc های دیتاست است.

“labels” : مشخص می‌کند که هر mfcc متعلق به کدام ژانر است.

به طور مثال داریم:

Dictionary = { “mapping” : [“blues” , “rock”] ,

“labels” : [0 , 1] ,

“mfcc”:[ […] , […] ]

}

سپس هر ترک را به چندین قسمت تقسیم میکنیم تا ترینینگ ست را افزایش دهیم.

111111111111111111111111111

اکنون از حلقه for و OS استفاده میکنیم تا تمام فولدرهای دیتا ست راه برویم.

ابتدا اسم فولدر ژانری که در آن هستیم را به دیکشنری قسمت mapping اضافه میکنیم روی تمام ترک هایی که درون پوشه ژانر مشخص شده با یک حلقه فور میچرخیم و همه ترک ها را با استفاده از librosia لود میکنیم. بعد لود هر ترک و به دست آوردن signal و samplerate هر ترک با همان librosia , برای هر ترک mfcc را حساب میکنیم به دیکشنری قسمتmfcc اضافه میکنیم و در قسمت lables دیکشینری مشخص میکنیم که به کدام ژانر متعلق است.

22222222222222222222222222

حال برای اطمینان چک میکنیم ابعاد Mfcc به دست آمده متناسب است سپس فایل جیسون را برای دیکشنری که از دیتا ست به دست آوردیم میسازیم.

333333333333333333333333

اکنون برای نمونه ما یک دیتا ست با 10 ژانر و برای هرکدام یک ترک موجود است را آماده کردیم و پس از ران کردن فایل جیسون را مشاهده خواهیم کرد

44444444