



۱. (۲۰٪) [پژوهشی]

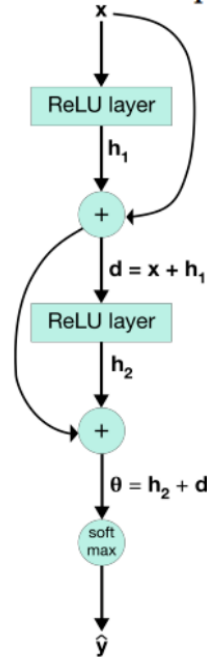
- ۱-۱) درباره اینکه چرا فاصله دو توزیع (مانند توزیع گاوس) از دو مجموعه داده متفاوت در یادگیری ماشین مهم است، تحقیق کنید و دو مثال از کاربرد آنها در روش‌های یادگیری ماشین بنزید و یک روش برای بدست آوردن فاصله میان دو توزیع نام ببرید.
- ۱-۲) حال، یک روش برای اینکه دو توزیع را به یکدیگر نزدیک کنیم، نام ببرید و شرح دهید.
- ۱-۳) تحقیق کنید منظور از auto machine learning چیست و یک روش آن در حوزه یادگیری عمیق را بیان کرده و به طور کلی و مختصر آن را شرح دهید.
- ۱-۴) درباره شبکه‌های Kolmogorov–Arnold Networks (KAN) که به تازگی آمده‌اند، و معروف شده‌اند، تحقیق کنید و نحوه کار آن را شرح دهید.

۲. (۲۰٪) [نظری: محو نشویم]

وقتی شبکه‌های عصبی بسیار عمیق می‌شوند (لایه‌های زیادی دارند)، آموزش دادن آنها به علت پدیده vanishing gradient مشکل می‌شود. ضرب‌های پشت هم می‌توانند موجب شوند مشتقات بسیار کوچک شوند و در نتیجه عملکرد شبکه بهبود نیابد یا حتی تضعیف شود. یک راه برای رفع این مشکل، استفاده از ResNet است که ایده اصلی آن استفاده از Skip Connection است که بیانگر پرش از یک یا چند لایه است (گراف زیر).



Neural network with skip connections



$$z_1 = W_1 x + b_1$$

$$h_1 = \text{ReLU}(z_1)$$

$$d = h_1 + x$$

$$z_2 = W_2 d + b_2$$

$$h_2 = \text{ReLU}(z_2)$$

$$\theta = h_2 + d$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = \text{CE}(y, \hat{y})$$

ابعاد متغیرها را به صورت $x \in \mathbb{R}^{D_x \times 1}$, $W_1 \in \mathbb{R}^{H \times D_x}$, $b_1 \in \mathbb{R}^H$, $W_2 \in \mathbb{R}^{D_y \times H}$, $b_2 \in \mathbb{R}^{D_y}$ و $\hat{y} \in \mathbb{R}^{D_y \times 1}$ در نظر بگیرید. همچنین فرض کنید که $D_x = D_y = H$ است. در این سوال قصد داریم $\frac{\partial J}{\partial x}$ را محاسبه کنیم. به ترتیب مراحل زیر را برای ایجاد جواب خود طی کنید:

الف) ابتدا $\delta_1 = \frac{\partial J}{\partial \theta}$ را بدست آورید.

ب) حال $\delta_2 = \frac{\partial J}{\partial z_2}$ را بدست آورید.

ج) $\delta_3 = \frac{\partial J}{\partial d}$ را محاسبه نمایید.

د) در نهایت با توجه به نتایج بخش‌های قبل $\frac{\partial J}{\partial x}$ را بدست آورید.

۳. (۳۰٪) [پیااده‌سازی: نجوا کرم یا نجوا گرم؟ 😞]

در این سوال گوهر بار، به دنبال کار با [whisper ai](https://whisper.ai) از شرکت open ai هستیم.

الف) یک جمله به صورت صوتی بگویید و در خروجی از این مدل، تشخیص این که چه زبانی است و متن آن را بگیرید و گزارش بدهید.



ب) این بار یک جمله به صورت صوتی بدهید و بخواهید به زبان چینی ترجمه کرده و نوشته آن را به شما بدهد باشد.

پ) یکی از روش‌های یادگیری انتقالی (transfer learning)، fine tune است، بدین شکل که مدل ما روی مجموعه داده‌ای آموزش داده شده است و تنها کاری که شما با مدل می‌کنید این است که آن مدل را با همان وزن‌ها می‌گیرید و فقط با مجموعه داده جدید آموزش می‌دهید تا روی مجموعه داده شما بهتر جواب دهد (و وزن‌ها با داده شما به‌روز شوند) و به گونه‌ای دانش مدل انتقال یافته و همچنین وفق داده شده است. در رابطه با مزیت‌های یادگیری انتقالی تحقیق کنید و دو روش دیگر از آن را نام ببرید و به صورت مختصر توضیح دهید.

ج) از این مجموعه داده استفاده کنید (۳ ساعت گفتار فارسی هر کدام ۱۰ ثانیه، در فایل اکسل رونویس هر فایل آمده است، نرخ نمونه برداری همان چیزی باشد که صدا با آن ذخیره شده است، `librosa.load(y=y, sr=None)`، حال ۸۰ درصد را برای آموزش (اگر نیاز داشتید از این ۸۰ درصد ۲۰ درصد را برای ارزیابی در نظر بگیرید) و ۲۰ درصد را برای تست انتخاب کنید، ابتدا این بیست درصد داده تست را به whisper بدهید (از مدل small استفاده شود) و میانگین WER: Word Error Rate (درصد کلمات اشتباه) آن را گزارش دهید، حال می‌خواهیم مدل را با این مجموعه داده fine tune کنیم. برای این کار مدل small از whisper را در نظر بگیرید، سپس آن را با این داده آموزش دهید. بسته به سخت‌افزار خود اگر می‌توانید ۵۰ epoch بگذارید. اگر سیستم قوی دارید بیشتر و اگر سیستم ضعیفی دارید هر چقدر کم‌تر باشد هم اشکال ندارد. حال روی همان داده تست خروجی از مدل بگیرید و میانگین WER را دوباره گزارش کنید. توجه شود که داده‌های تست در هر دو حالت باید کاملاً یکسان باشند.

۴. (۳۰٪ + ۲۰٪ امتیازی) [پایاده‌سازی: شبکه LSTM]

در این تمرین بنا داریم تا ژانرهای موسیقیایی را با طراحی شبکه‌های LSTM (در تمرین قبلی از شبکه CNN استفاده کردیم) از هم تشخیص دهیم.

(<https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification>)



که ۱۰ کلاس ژانر موسیقی در آن است. همه این صداها، طول ۳۰ ثانیه دارد. برای پیاده‌سازی شبکه LSTM می‌توانید از کتابخانه‌ها و ابزارهای دلخواه بهره ببرید.

گام ۱- فراخوانی دادگان و بخش‌بندی 80 / 20 آن (train / val): این داده، 10 پوشه دارد که هر کدام بیانگر یک ژانر (کلاس) است و در هر پوشه، 100 فایل صدا است. لازم است که تمامی صداها را در کنار هم داشته باشید و سپس 80 / 20 را اعمال کنید.

گام ۲- با در نظر گرفتن طول فریم ۲۵ میلی ثانیه و ۱۳ ویژگی mfcc، ویژگی‌های mfcc را استخراج کنید.

گام ۳- ساخت و آموزش مدل: ویژگی دادگان را به عنوان ورودی ابتدا به یک لایه LSTM با ۶۴ hidden state دهید، سپس خروجی آن را دوباره به یک لایه دیگر با 64 hidden state دهید، حال به یک لایه dense با ۶۴ نورون و تابع فعال‌ساز relu دهید، سپس یک لایه drop out به اندازه ۰.۳ بگذارید و در آخر یک لایه dense با ۱۰ نورون خروجی و تابع فعال‌ساز softmax بگذارید. از بهینه‌ساز adam با learning rate = 0.0001 استفاده کرده و همچنین تابع loss را sparse_categorical_crossentropy در نظر بگیرید، مدل را با ۳۰ epoch و batch size برابر با ۳۲ آموزش دهید.

در هر epoch دقت را برای آموزش و ارزیابی گزارش دهید.

به نظر شما دلایل کم بودن یا زیاد بودن دقت روی این دادگان با این روش چیست؟

کارایی این روش را با تمرین قبلی (شبکه CNN) مقایسه کنید.

بخش امتیازی

الف) به نظر شما چرا به جای دو لایه LSTM با hidden state 64، از یک لایه LSTM با hidden state ۱۲۸ استفاده نشده؟ یک بار این گونه مدل را با تمام فرضیات قبل آموزش دهید و گزارش دهید.

ب) گام‌های بیان شده در تمرین را با Bidirectional LSTM تکرار کنید و نتیجه را گزارش کنید. کارایی شبکه یک طرفه و دوطرفه را مقایسه کنید.