



دانشگاه تهران

دانشکده سامانه‌های هوشمند

یادگیری ماشین

تمرین پنجم

استاد درس

دکتر سامان هراتی‌زاده

زمان تحویل: ۲۱ دی

پاییز ۱۴۰۲

## یادگیری ماشین

### تمرین «۵»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۸

دستیاران آموزشی

کامیار رحمانی

بهناز ریوندی

امیرمحمد کویش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

### فهرست

۱. نظری- تخمین بیشینه شباهت ..... ۱
۲. نظری- الگوریتم EM ..... ۱
۳. نظری- روش گشتاورها ..... ۲
۴. پیاده سازی- مدل تشخیص گفتار با استفاده از پرسپترون چند لایه ..... ۲
۵. نظری- خوشه بندی ..... ۷
۶. نظری- خوشه بندی K-means ..... ۷

### شکل ها

- شکل ۱ معماری پرسپترون چند لایه ..... ۳
- شکل ۲ روند اجرای کد ..... ۳
- شکل ۳ حالت های واج برای واج مشخص ..... ۴
- شکل ۴ استخراج ویژگی ..... ۵
- شکل ۵ نمایش داده ..... ۶
- شکل ۶ حرکت طیف از فریم های مجاور ..... ۶

# یادگیری ماشین

## تمرین «۵»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۸

دستیاران آموزشی

کامیار رحمانی

بهناز ریوندی

امیرمحمد کویش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

۱. (۱۰٪) [نظری-تخمین بیشینه شباهت]<sup>۱</sup> توزیع احتمال Pareto در اقتصاد کاربرد زیادی دارد. رابطه این توزیع به صورت زیر است:

$$P(x) = \frac{\theta b^\theta}{x^{\theta+1}}$$

که در رابطه فوق  $\theta$  و  $b$  پارامترهای مدل هستند. فرض کنید نمونه های  $D = \{x_1, \dots, x_n\}$  به صورت i.i.d از توزیع احتمال گفته شده آمده باشند.

۱. تابع log-likelihood را تشکیل دهید و تخمین گر maximum likelihood را برای پارامتر  $\theta$  به دست آورید.

۲. توزیع احتمال پیشین زیر را برای پارامتر  $\theta$  در نظر بگیرید.

$$p(\theta) = \text{Gamma}(\theta | \alpha, \beta) = c \theta^{\alpha-1} e^{-\beta\theta}$$

که در رابطه بالا  $c$  یک ضریب ثابت و  $\alpha, \beta$  پارامترهای توزیع گاما هستند. توزیع احتمال پسین را برای پارامترهای  $\theta$  بدست آورید (راهنمایی:  $a^b = e^{b \ln a}$ )

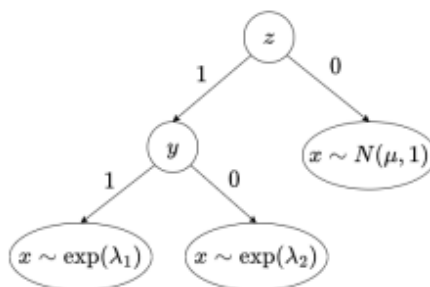
$$p(\theta|D) = ?$$

۳. آیا توزیع احتمال پیشین فوق یک conjugate prior برای پارامتر  $\theta$  است؟ توضیح دهید.

۴. با استفاده از توزیع احتمال پیشین فوق، تخمین گر MAP برای پارامتر  $\theta$  چیست؟ (راهنمایی: مقدار بیشینه توزیع گاما در نقطه  $\theta = \frac{\alpha-1}{\beta}$  رخ می دهد).

۵. آیا اگر  $n \rightarrow \infty$  آنگاه تخمین گر MAP به تخمین گر ML میل می کند؟ چرا؟

۲. (۱۵٪) [نظری-الگوریتم EM] فرض کنید متغیر تصادفی  $x \in R$  مطابق درخت زیر تولید می شود:



که در شکل بالا  $z, y$  متغیرهای پائینی و مستقل از یک دیگر هستند. اگر احتمال ۱ بودن متغیرهای  $z, y$  به ترتیب برابر با  $\alpha, \beta$  باشد. آنگاه توزیع احتمال توام هر سه متغیر  $x, y, z$  به صورت زیر خواهد بود:

$$P(x, y, z) = \left( (1-\alpha) \frac{1}{2\pi} \exp\left(-\frac{(x-\mu)^2}{2}\right) \right)^{1-z} \left( \alpha ((1-\beta)\lambda_2 \exp(-\lambda_2 x))^{1-y} (\beta\lambda_1 \exp(-\lambda_1 x))^y \right)^z$$

فرض کنید مجموعه داده های  $D = \{x_1, \dots, x_n\}$  را که به صورت i.i.d هستند، در اختیار داریم. حال می خواهیم با استفاده از روش EM، پارامترهای توزیع فوق را تخمین بزنیم.

۱. تابع log-likelihood را تشکیل دهید.

۲. امید ریاضی تابع log-likelihood را نسبت به متغیرهای پنهان (یعنی  $y_i, z_i$ ) بدست آورید. توجه داشته باشید باید به داده مشاهده شده یعنی  $x_i$  این کار را انجام دهید (راهنمایی:  $E[y_i z_i | x_i] = P(y_i z_i = 1 | x_i)$ ).

<sup>1</sup> Maximum Likelihood Estimation

## یادگیری ماشین

### تمرین «۵»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۸

دستیاران آموزشی

کامیار رحمانی

بهناز ریوندی

امیرمحمد کویش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

۳. با استفاده از امید ریاضی بدست آمده از قسمت قبل، مقدار بهینه پارامترهای  $\beta, \mu, \lambda_1$  را بدست آورید.

۳. (۱۰٪) [نظری-روش گشتاورها]<sup>۱</sup> فرض کنید  $D = \{x_1, \dots, x_n\}$  نمونه‌هایی از توزیع زیر باشند:

$$p(x|\theta) = \frac{1}{\theta} x^{\frac{1-\theta}{\theta}}$$

۱. تخمین گر maximum likelihood را برای توزیع فوق تشکیل دهید و پارامتر  $\theta$  را تخمین بزنید.

۲. با استفاده از Method of Moments تخمینی برای پارامتر  $\theta$  بدست آورید.

۴. (۴۰٪) [پایه سازی-مدل تشخیص گفتار با استفاده از پرسپترون چند لایه] هدف اصلی این تمرین بررسی شبکه‌های

عصبی برای تشخیص گفتار است.

پس از حل این تمرین شما یاد خواهید گرفت: چگونه با استفاده از پرسپترون چند لایه<sup>۲</sup> مساله‌های طبقه‌بندی را حل کنید: - نحوه پیاده سازی پرسپترون چند لایه، - نحوه مدیریت داده‌ها، - نحوه آموزش مدل، - نحوه بهینه سازی مدل، و به بررسی فراپارامترها برای بهینه سازی مدل بپردازید: - شناسایی و جدول بندی تمام انتخاب‌های مختلف طراحی/معماری و فراپارامترها، - شناسای راه‌کارهایی برای جستجو در فضای جواب‌ها برای یافتن بهترین جواب ممکن

گفتار برای بشر طبیعی‌ترین و کارآمدترین ابزار مبادله اطلاعات است. تشخیص گفتار زیرشاخه‌ای از زبان‌شناسی محاسباتی می‌باشد. این زیرشاخه با تکنولوژی‌هایی کار می‌کند که داده‌های صوتی (گفتار) را به‌عنوان ورودی دریافت و تجزیه و تحلیل می‌کنند. در این رویکرد گفتار به کمک تعدادی واحد آوایی (مانند کلمه، هجا، سه واجی یا واج) مدل می‌شود و برای بازشناسی نیز از تشخیص این واحدها و کنار هم قرار دادن آنها، متن متناسب با گفتار تشخیص داده می‌شود. بنابراین سیستم تشخیص گفتار نوعی فناوری است که به یک رایانه این امکان را می‌دهد که گفتار و کلمات گوینده را بازشناسی و خروجی آن را به قالب مورد نظر، مانند «متن»، ارائه کند. داده‌های صوتی یا گفتاری<sup>۳</sup> شامل ضبط صدا، از جمله گفتار، موسیقی یا سایر سیگنال‌های صوتی است.

کوچک‌ترین بخش گفتار واج می‌باشد که به کمک آن تکواژ ساخته می‌شود و یک یا چند تکواژ یک واژه را می‌سازند. برای مثال واژه «ما» از دو واج /م/ (m/) و /ا/ (v/) تشکیل شده است بخش عمده واج، جداسازی واحدهای گفتاری از یکدیگر و ایجاد تمایز بین معانی واحدهای گفتاری است و می‌توان گفت از ترکیب واج‌ها و به عبارت دیگر، «از ترکیب واحدها و قالب‌های صوتی با واحدها و قالب‌های معنایی»، سامانه ارتباط یعنی زبان به وجود می‌آید.

مجموعه داده ارائه شده در این تمرین شامل داده‌های گفتاری به شکل Mel spectrograms است. (در قسمت بعدی به طور کامل توضیح داده می‌شود) داده‌های آموزشی شامل واج‌های متناظر برای این داده‌ها هستند و ما در این تمرین پرسپترون چندلایه‌ای می‌سازیم که می‌تواند حالت‌های واج را در داده‌های آموزشی تشخیص دهد و برچسب‌گذاری کند. پرسپترون چندلایه نوعی شبکه عصبی است که از چندین لایه پرسپترون تشکیل شده است که ویژگی‌ها و الگوهای داده‌ها را یاد می‌گیرد. پس از تکمیل تمرین شما مهارت کافی برای پیاده سازی شبکه های عصبی و علاوه بر آن، تنظیم و تغییر پارامترها برای رسیدن به جواب بهینه را بدست می‌آورید.

<sup>1</sup> Method of Moments (MOM)

<sup>2</sup> MLP

<sup>3</sup> Speech

# یادگیری ماشین

## تمرین «۵»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۸

دستیاران آموزشی

کامیار رحمانی

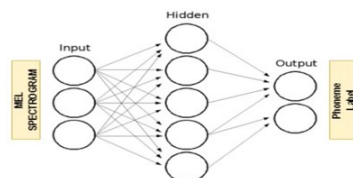
بهناز ریوندی

امیرمحمد کویش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲



شکل ۱ معماری پرسپترون چند لایه

### ۱. مقدمه: طبقه بندی در سطح فریم داده های صوتی

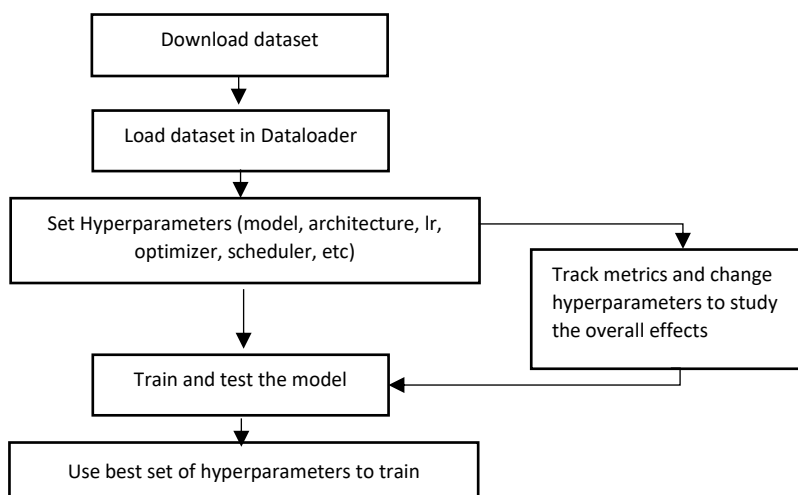
در این تمرین از شبکه های عصبی در زمینه ی تشخیص گفتار استفاده می کنیم. مجموعه داده ای از ضبط های صوتی (گفته ها) و برچسب های حالت واج (فرعی) آنها در اختیار شما قرار می گیرد که این داده ها از مقالات منتشر شده در وال استریت ژورنال (WSJ) که به شکل صوت در آمده می باشد و با استفاده از متن اصلی برچسب گذاری شده است. اگر قبلاً با داده های گفتاری مواجه نشده اید یا نام واج ها یا طیف نگارها را نشنیده اید، در قسمت های بعدی بیشتر توضیح خواهیم داد.

### آشنایی با مجموعه دادگان

داده های آموزشی شامل موارد زیر است: داده های صوتی<sup>۱</sup>، برچسب های حالت واج برای هر فریم داده های آزمون شامل موارد زیر است: داده های صوتی، برچسب های حالت واج داده نشده است

### ۲. راهنمایی

به همراه صورت سوال یک فایل جویپتر نوت بوک<sup>۲</sup> در اختیار شما قرار گرفته است که با توجه به توضیحات داده شده باید آن را تکمیل نمایید. این فایل شامل بلوک های اصلی برای آموزش پرسپترون چندلایه ای است. برای اطلاعات بیشتر در مورد روند اجرای این تمرین تصویر زیر را مشاهده کنید.



شکل ۲ روند اجرای کد

### ۳. آشنایی با واج ها و حالت های واج

همانطور که حروف عناصر تشکیل دهنده زبان نوشتاری می باشند واج ها نیز عناصر تشکیل دهنده گفتار هستند. سیستم های تشخیص گفتار با بکار گیری روش های مختلف طبقه بندی و شناسایی الگو قادر به تشخیص واژگان هستند که البته برای افزایش دقت در شناسایی

<sup>۱</sup> Raw Mel Spectrogram Frames

<sup>۲</sup> IPython Notebook (.ipynb)

## یادگیری ماشین

### تمرین «۵»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۸

دستیاران آموزشی

کامیار رحمانی

بهناز ریوندی

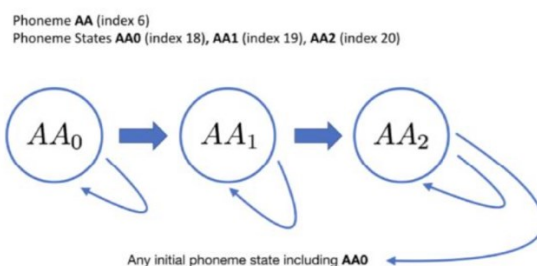
امیرمحمد کویش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

از یک فرهنگ لغات نیز در انتهای سیستم استفاده می‌شود. دو مدل مسلط در این حوزه مدل مخفی مارکوف<sup>۱</sup> و مدل شبکه عصبی<sup>۲</sup> هستند. این روش‌ها اساساً برای مشخص کردن اطلاعات پنهان از سیستم، از اطلاعاتی که برای سیستم شناخته شده هستند استفاده می‌کنند. مدل Hidden Markov رایج‌ترین مدل است که احتمال داده‌های گفتاری مشاهده شده را به حداکثر می‌رساند. در این مدل واج مانند یک پیوند در یک زنجیره است و هنگامی این زنجیره تکمیل می‌شود، یک کلمه بوجود می‌آید. در این مدل برای هر واج، ۳ حالت واجی مربوطه وجود دارد. نمودار انتقال حالات واج برای یک واج معین به شرح زیر نشان داده شده است:



شکل ۳ حالت‌های واج برای واج مشخص

برای این تمرین در مجموع ۴۰ واج برای زبان در نظر خواهیم گرفت.

["+BREATH+", "+COUGH+", "+NOISE+", "+SMACK+", "+UH+", "+UM+", "AA", "AE", "AH", "AO", "AW", "AY", "B", "CH", "D", "DH", "EH", "ER", "EY", "F", "G", "HH", "IH", "IY", "JH", "K", "L", "M", "N", "NG", "OW", "OY", "P", "R", "S", "SH", "SIL", "T", "TH", "UH", "UW", "V", "W", "Y", "Z", "ZH"]

در این تمرین شما با استفاده از شبکه عصبی یک فریم (به‌علاوه بردار زمینه اختیاری) را به عنوان ورودی دریافت می‌کنید و احتمالات کلاس را برای همه ۴۰ حالت خروجی تولید می‌کنید.

#### ۴. استخراج ویژگی‌ها

گام بعدی، استخراج ویژگی‌هایی است که برای آموزش دادن مدل به آن‌ها نیاز است. به طور کلی ویژگی‌های سیگنال صوت (در حوزه سیگنال) می‌توانند در سه دسته‌ی زیر قرار گیرند:

۱. ویژگی‌های حوزه زمان<sup>۳</sup>: این ویژگی‌های قابل استخراج از شکل موج در حوزه زمان<sup>۴</sup> می‌باشند. برای مثال: amplitude envelope, root-mean square energy, zero crossing rate, ... بدیهی است که فرکانس نیز تا حد زیادی می‌تواند توصیف‌کننده سیگنال صوت باشد. بنابراین دسته دیگری از ویژگی‌ها نیز مورد نیاز می‌باشند.

۲. ویژگی‌های حوزه فرکانس<sup>۵</sup>: می‌توان با اعمال تبدیل فوریه بر روی سیگنال در حوزه زمان، سیگنال را در حوزه فرکانس نمایش داد. در این صورت برخی از ویژگی‌هایی که از این نمایش قابل استخراج می‌باشند عبارتند از: Spectral centroid, Band energy ratio, Spectral flux, ...

همانطور که گفته شد، در هر یک از حوزه‌های بالا، تنها ویژگی‌های مربوط به همان حوزه قابل استخراج است. اما دسته‌ی دیگری از ویژگی‌ها وجود دارند که اطلاعاتی را در هر دو این حوزه‌ها در اختیار ما قرار می‌دهند.

۳. ویژگی‌های حوزه زمان-فرکانس<sup>۶</sup>: برای استخراج اینگونه ویژگی‌ها می‌توان از نمایش‌های حوزه زمان-فرکانس استفاده نمود: Mel-Spectrogram, Constatn-Q transform, Spectrogram, ...

<sup>1</sup> Hidden Markov Model (HMM)

<sup>2</sup> Neural Network Model (NN)

<sup>3</sup> Time domain features

<sup>4</sup> Waveform

<sup>5</sup> Frequency domain features

<sup>6</sup> Time-frequency features

## یادگیری ماشین

### تمرین «۵»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۸

دستیاران آموزشی

کامیار رحمانی

بهناز ریوندی

امیرمحمد کویش پور

دکتر سامان هراتی زاده

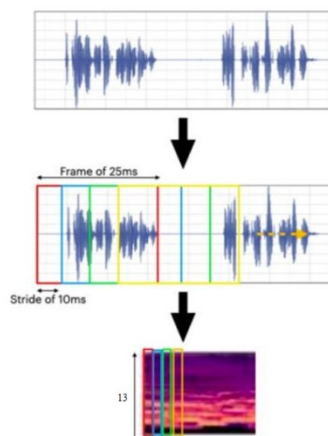
دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

«طیف‌سنج‌ها» (Spectrograms) روش‌های مفیدی برای بصری‌سازی طیف فرکانس‌های یک صدا و چگونگی تغییر آن‌ها در طول یک بازه زمانی هستند. اسپکتروگرام از روی خروجی تبدیل فوری به زمان کوتاه بدست می‌آید. برای بدست آوردن اسپکتروگرام در ابتدا تبدیل فوری به زمان کوتاه روی سیگنال اعمال می‌شود، (سیگنال به چندین بازه زمانی تبدیل شده و روی هر کدام یک تبدیل فوری جدا اعمال می‌شود) و در نتیجه آن به ازای هر بازه زمانی یک طیف فرکانسی بدست می‌آید. سپس ضرایب هر کدام از طیف‌های فرکانسی براساس میزان دامنه ای که دارند به یک کد رنگی تبدیل می‌شوند. که در نتیجه آن ما یک نقشه رنگی برحسب زمان و فرکانس بدست می‌آید که ما می‌توانیم با تحلیل آن متوجه شویم که طیف فرکانسی سیگنال در طول زمان به چه صورت تغییر می‌کند.

در این تمرین برای هر فایل صوتی در مجموعه داده، یک نمایش melspectrogram استخراج شده و در اختیار شما قرار گرفته است. نکته: تبدیل فوری کوتاه مدت بر روی بخش‌های کوچک شکل موج که فریم نامیده می‌شوند، هر کدام ۲۵ میلی‌ثانیه می‌باشد، انجام شده است. در نتیجه‌ی این تبدیل یک بردار واحد تولید می‌شود. از آنجایی که ما بین هر فریم stride حدود ۱۰ میلی‌ثانیه استفاده می‌کنیم، در نهایت به ۱۰۰ بردار در ثانیه داده می‌رسیم. که هر بردار یک بردار ۲۸ بعدی ویژگی می‌باشد (برای جزئیات دقیق نحوه انجام این کار، به پیوندهای موجود در بخش پیوست مراجعه کنید). برای یک داده صوتی T ثانیه‌ای، ماتریسی با ابعاد  $(T \times 100, 28)$  تولید می‌شود. توجه داشته باشید که در مجموعه داده‌ای که در اختیار شما قرار داده شده است، تمامی این پیش پردازش‌ها انجام شده است و ملسپکتروگرام‌های شکل نهایی  $(28, *)$  در اختیار شما قرار گرفته است.

جمع بندی: داده‌های ارائه شده در این بخش شامل این ملسپکتروگرام‌ها و برچسب‌های واجی برای هر بردار ۲۸ بعدی در ملسپکتروگرام است و هدف پیش بینی برچسب بردار ۲۸ بعدی خاص در یک فریم داده صوتی می‌باشد.



شکل ۴ استخراج ویژگی

## یادگیری ماشین

### تمرین «۵»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۸

دستیاران آموزشی

کامیار رحمانی

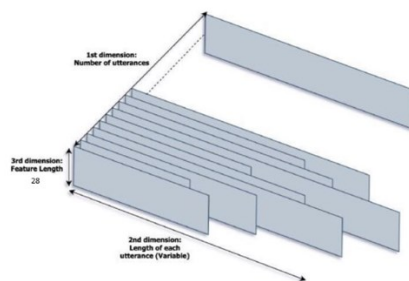
بهناز ریوندی

امیرمحمد کویش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲



شکل ۵ نمایش داده

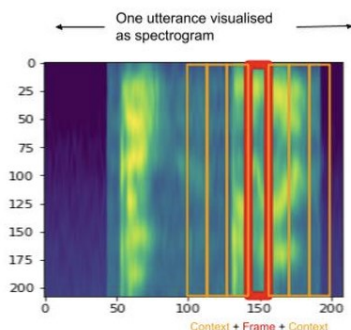
#### ۵. بردار زمینه

با توجه به اینکه هر داده صوتی تنها ۲۵ میلی‌ثانیه را شامل می‌شود و ممکن است این بردار ویژگی به تنهایی کافی نباشد، راه حلی پیشنهاد دهید. (راه حل: راه حلی که می‌توانیم استفاده کنیم این است که برداری به نام «بردار زمینه» به اندازه  $K$  در اطراف هر بردار اضافه کنیم.) برای مثال، بردار زمینه با اندازه ۵ به این معنی می‌باشد که ما یک ورودی با اندازه (۱۱، ۲۸) به شبکه ارائه می‌دهیم - اندازه ۱۱ ممکن است به صورت زیر توضیح داده شود: بردار برای پیش‌بینی برچسب از اطلاعات ۵ بردار قبل از این بردار، و ۵ بردار بعد از آن می‌تواند استفاده کند.

راه‌های مختلفی برای پیاده‌سازی این کار وجود دارد، راه حلی دیگر شما پیشنهاد دهید

**توجه:** ممکن است بخواهید برای نتایج بهتر، مقداری padding صفر به هر داده اضافه کنید. به عنوان مثال، اگر یک داده صوتی منفرد با بعد (۲۸، ۱۰۰۰) را در نظر بگیریم و زمینه را ۵ در نظر بگیریم، می‌خواهیم قبل و بعد از این نمونه، padding صفر را اضافه کنیم تا به بعد (۲۸، ۱۰۱۰) تبدیل شود. به همین ترتیب لایه ورودی  $(2+1) \times \text{اندازه زمینه} \times 28$  گره خواهد داشت. (چرا؟)

بردار زمینه یک فرآیند ساده‌تر است و مقدار پیشنهادی بردار زمینه برای تنظیم این تمرین بین ۵۰-۰ است. برای اطلاعات بیشتر در مورد بردار زمینه، به پیوست مراجعه کنید



شکل ۶ حرکت طیف از فریم‌های مجاور

#### ۶. تنظیم پارامترهای مدل

در یادگیری عمیق، هایپرپارامترها شامل متغیرهایی هستند که برای تنظیم شبکه عصبی استفاده می‌شوند، در زیر چند تغییر در پارامترها وجود دارد که ممکن است به شما کمک کند.

Hyperparameters	Values
Number of Layers	2-8
Activations	ReLU, LeakyReLU, softplus, tanh, sigmoid



# یادگیری ماشین

## تمرین «۵»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۸

دستیاران آموزشی

[کامیار رحمانی](#)

[بهناز ریوندی](#)

[امیرمحمد کویش پور](#)

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

Batch Size	64, 128, 256, 512, 1024, 2048
Architecture	Cylinder, Pyramid, Inverse-Pyramid, Diamond
Dropout	0-0.5, Dropout in alternate layers
LR Scheduler	Fixed, StepLR, ReduceLROnPlateau, Exponential, CosineAnnealing
Weight Initialization	Gaussian, Xavier, Kaiming(Normal and Uniform), Random, Uniform
Context	0-50
Batch-Norm	Before or After Activation, Every layer or Alternate Layer or No Layer
Optimizer	Vanilla SGD, Nesterov's momentum, RMSProp, Adam
Regularization	Weight Decay
LR	0.001, you can experiment with this
Normalization	You can try Cepstral Normalization

### معماری اولیه پیشنهادی برای شروع

بردار زمینه: ۲۰، تعداد لایه پنهان: ۰، تابع فعال سازی: RELU، میزان یادگیری: ۰-۳، بهینه ساز: Adam

برای اطلاعات بیشتر در مورد این سوال، لطفا فایل ضمیمه [appendix] آماده شده را مطالعه کنید.

۵. (۱۰٪) [نظری-خوشه بندی] به سوال های پرسیده شده، به طور کامل و در صورت نیاز به تعداد گام خواسته شده الگوریتم ها را اجرا کنید.
۱. انتظار دارید در الگوریتم DBSCAN افزایش یا کاهش اندازه شعاع همسایگی  $\epsilon$  (بدون تغییر مقدار MinPts) چگونه بر تعداد خوشه ها و تعداد نمونه هایی که نويز تشخیص داده می شوند، اثر بگذارد؟ چرا؟
۲. نمونه های زیر که توسط یک ویژگی  $X$  توصیف شده اند را به روش سلسله مراتبی پایین به بالا (با استفاده از معیار فاصله Single Linkage) به دو خوشه تقسیم کنید و کام های خوشه بندی را بنویسید و معیار silhouette را برای نقطه  $x=2$  محاسبه کنید.

X
5
2
6.3
8
4.5
1
7.1

### ۶. (۱۵٪) [نظری-خوشه بندی K-means]

۱. ثابت کنید الگوریتم k-means همگرا می شود.
۲. برای دادگانی با  $n$  داده و  $k$  خوشه ( $k > 2$ )، نصف داده ها در ناحیه متمرکز و نصف دیگر در ناحیه ای با چگالی کمتر قرار گرفته اند و این دو ناحیه تقریباً جدا از هم اند، پس از خوشه بندی دادگان (با احتساب MSE) آیا در نهایت مرکز خوشه ها به صورت یکنواخت بین دو ناحیه ذکر شده توزیع می شوند؟ یا در ناحیه ای مراکز تجمع بیشتری دارند؟ کدام ناحیه؟
۳. الگوریتم برگرفته از k-means را اینگونه در نظر بگیرید که در اولین مرحله انتخاب مراکزها، اولین مرکز به صورت تصادفی انتخاب شود ولی مرکز بعدی نقطه ای با بیشترین احتمال به دست آمده از فرمول زیر باشد. مراکز بعدی نیز به همین صورت. در انتخاب مرکز در مراحل بعدی (هنگام به روز کردن مراکزها) نیز نقطه با بیشترین احتمال فرمول زیر را به عنوان مرکز جدید بر می گزینیم.

$$\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$$

$D(x)$  طول کوتاهترین فاصله بین داده  $x$  تا نزدیک ترین مرکز است که قبلاً انتخاب شده است.  $X$  نیز مجموعه تمام داده ها است.

نتیجه خوشه بندی این الگوریتم را نسبت به نتیجه k-means مقایسه کنید؟ از لحاظ سرعت همگرایی نیز مقایسه کنید؟

## یادگیری ماشین

### تمرین «۵»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۸

دستیاران آموزشی

کامیار رحمانی

بهناز ریوندی

امیرمحمد کویش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

۴. آیا استفاده از معیار فاصله کوسینوسی در k-means برای داده های شامل ویژگی زمان (Time Series Data) معیار مناسبی است؟ توضیح دهید؟ و اگر مناسب نیست، چه معیار اندازه گیری فاصله ای مناسب تر است؟