



دانشگاه تهران

دانشکده سامانه‌های هوشمند

یادگیری ماشین

تمرین پنجم

استاد درس

دکتر سامان هراتی‌زاده

زمان تحویل: ۱۸ 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 ۲
- شکل ۲ روند اجرای کد ۳
- شکل ۳ حالت های واج برای واج مشخص ۴
- شکل ۴ استخراج ویژگی ۵
- شکل ۵ نمایش داده ۵
- 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]^۱ 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 θ و β 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 λ به دست آورید.

۲. توزیع احتمال پیشین زیر را برای پارامتر λ در نظر بگیرید.

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

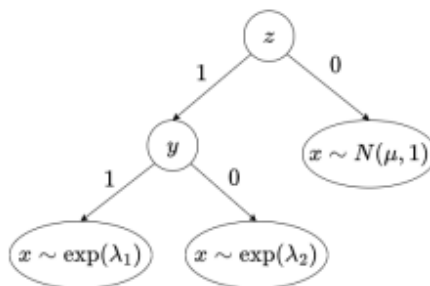
که در رابطه بالا c یک ضریب ثابت و α, β پارامترهای توزیع گاما هستند. توزیع احتمال پسین را برای پارامترهای θ بدست آورید (راهنمایی: $a^b = e^{b \ln a}$)

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

۴. با استفاده از توزیع احتمال پیشین فوق، تخمین گر MAP برای پارامتر θ 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 باینری و مستقل از یک دیگر هستند. اگر احتمال ۱ بودن متغیرهای z, y به ترتیب برابر با α, β 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: $E[y_i z_i | x_i] = P(y_i z_i = 1 | x_i)$).

۳. با استفاده از امید ریاضی بدست آمده از قسمت قبل، مقدار بهینه پارامترهای β, μ, λ_1 را بدست آورید.

¹ 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 سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

۳. (۱۰٪) [نظری-روش گشتاورها]^۱ فرض کنید $D = \{x_1, \dots, x_n\}$ نمونه هایی 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 θ را تخمین بزنید.

۲. با استفاده از Method of Moments تخمینی برای پارامتر θ بدست آورید.

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

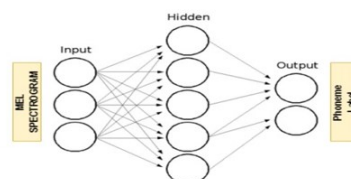
عصبی برای تشخیص گفتار است.

پس از حل این تمرین شما یاد خواهید گرفت: چگونه با استفاده از پرسپترون چند لایه^۲ مساله های طبقه بندی را حل کنید: - نحوه پیاده سازی پرسپترون چند لایه، - نحوه مدیریت داده ها، - نحوه آموزش مدل، - نحوه بهینه سازی مدل، و به بررسی فرآیندها برای بهینه سازی مدل بپردازید: - شناسایی و جدول بندی تمام انتخاب های مختلف طراحی/معماری و فرآیندها، - شناسایی راه کارهایی برای جستجو در فضای جواب ها برای یافتن بهترین جواب ممکن

گفتار برای بشر طبیعی ترین و کارآمدترین ابزار مبادله اطلاعات است. تشخیص گفتار زیرشاخه ای از زبان شناسی محاسباتی می باشد. این زیرشاخه با تکنولوژی هایی کار می کند که داده های صوتی (گفتار) را به عنوان ورودی دریافت و تجزیه و تحلیل می کنند. در این رویکرد گفتار به کمک تعدادی واحد آوایی (مانند کلمه، هجا، سه واجی یا واج) مدل می شود و برای بازشناسی نیز از تشخیص این واحدها و کنار هم قرار دادن آنها، متن متناسب با گفتار تشخیص داده می شود. بنابراین سیستم تشخیص گفتار نوعی فناوری است که به یک رایانه این امکان را می دهد که گفتار و کلمات گوینده را بازشناسی و خروجی آن را به قالب مورد نظر، مانند «متن»، ارائه کند. داده های صوتی یا گفتاری^۳ شامل ضبط صدا، از جمله گفتار، موسیقی یا سایر سیگنال های صوتی است.

کوچک ترین بخش گفتار واج می باشد که به کمک آن تکواژ ساخته می شود و یک یا چند تکواژ یک واژه را می سازند. برای مثال واژه «ما» از دو واج /م/ و /ن/ تشکیل شده است بخش عمده واج، جداسازی واحدهای گفتاری از یکدیگر و ایجاد تمایز بین معانی واحدهای گفتاری است و می توان گفت از ترکیب واج ها و به عبارت دیگر، «از ترکیب واحدها و قالب های صوتی با واحدها و قالب های معنایی»، سامانه ارتباط یعنی زبان به وجود می آید.

مجموعه داده ارائه شده در این تمرین شامل داده های گفتاری به شکل Mel spectrograms است. (در قسمت بعدی به طور کامل توضیح داده می شود) داده های آموزشی شامل واج های متناظر برای این داده ها هستند و ما در این تمرین پرسپترون چندلایه ای می سازیم که می تواند حالت های واج را در داده های آموزشی تشخیص دهد و برچسب گذاری کند. پرسپترون چندلایه نوعی شبکه عصبی است که از چندین لایه پرسپترون تشکیل شده است که ویژگی ها و الگوهای داده ها را یاد می گیرد. 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 می آورید.



شکل ۱ معماری پرسپترون چند لایه

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

¹ Method of Moments (MOM)

² MLP

³ 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 سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

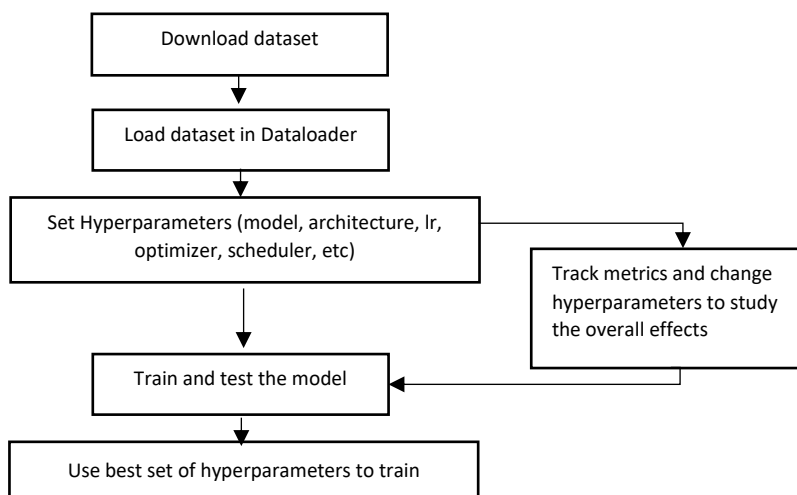
در این تمرین از شبکه‌های عصبی در زمینه‌ی تشخیص گفتار استفاده می‌کنیم. مجموعه داده‌ای از ضبط‌های صوتی (گفته‌ها) و برچسب‌های حالت واج (فرعی) آنها در اختیار شما قرار می‌گیرد که این داده‌ها از مقالات منتشر شده در وال استریت ژورنال (WSJ) که به شکل صوت در آمده می‌باشد و با استفاده از متن اصلی برچسب گذاری شده است. اگر قبلاً با داده‌های گفتاری مواجه نشده‌اید یا نام واج‌ها یا طیف‌نگارها را نشنیده‌اید، در قسمت‌های بعدی بیشتر توضیح خواهیم داد.

آشنایی با مجموعه دادگان

داده‌های آموزشی شامل موارد زیر است: داده‌های صوتی^۱، برچسب‌های حالت واج برای هر فریم داده‌های آزمون شامل موارد زیر است: داده‌های صوتی، برچسب‌های حالت واج داده نشده است

۲. راهنمایی

به همراه صورت سوال یک فایل جویپتر نوت‌بوک^۲ در اختیار شما قرار گرفته است که با توجه به توضیحات داده شده باید آن را تکمیل نمایید. این فایل شامل بلوک‌های اصلی برای آموزش پرسپترون چندلایه‌ای است. برای اطلاعات بیشتر در مورد روند اجرای این تمرین تصویر زیر را مشاهده کنید.



شکل ۲ روند اجرای کد

۳. آشنایی با واج‌ها و حالت‌های واج

همانطور که حروف عناصر تشکیل دهنده زبان نوشتاری می‌باشند واج‌ها نیز عناصر تشکیل دهنده گفتار هستند. سیستم‌های تشخیص گفتار با بکارگیری روش‌های مختلف طبقه‌بندی و شناسایی الگو قادر به تشخیص واژگان هستند که البته برای افزایش دقت در شناسایی از یک فرهنگ لغات نیز در انتهای سیستم استفاده می‌شود. دو مدل مسلط در این حوزه مدل مخفی مارکوف^۳ و مدل شبکه عصبی^۴ 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:

^۱ Raw Mel Spectrogram Frames

^۲ IPython Notebook (.ipynb)

^۳ Hidden Markov Model (HMM)

^۴ Neural Network Model (NN)

یادگیری ماشین

تمرین «۵»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۸

دستیاران آموزشی

کامیار رحمانی

بهناز ریوندی

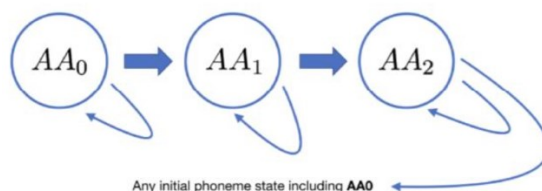
امیرمحمد کویش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

Phoneme AA (index 6)
Phoneme States AA0 (index 18), AA1 (index 19), AA2 (index 20)



شکل ۳ حالت های 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 (به علاوه بردار زمینه اختیاری) را به عنوان ورودی دریافت می کنید و احتمالات کلاس را برای همه ۴۰ حالت خروجی تولید می کنید.

۴. استخراج ویژگی ها

گام بعدی، استخراج ویژگی هایی است که برای آموزش دادن مدل به آن ها نیاز است. به طور کلی ویژگی های سیگنال صوت (در حوزه سیگنال) می توانند در سه دسته ی زیر قرار گیرند:

۱. ویژگی های حوزه زمان^۱: این ویژگی های قابل استخراج از شکل موج در حوزه زمان^۲ می باشند. برای مثال: amplitude envelope, root-mean square energy, zero crossing rate, ...

بدیهی است که فرکانس نیز تا حد زیادی می تواند توصیف کننده سیگنال صوت باشد. بنابراین دسته دیگری از ویژگی ها نیز مورد نیاز می باشند.

۲. ویژگی های حوزه فرکانس^۳: می توان با اعمال تبدیل فوریه بر روی سیگنال در حوزه ی زمان، سیگنال را در حوزه ی فرکانس نمایش داد. در این صورت برخی از ویژگی هایی که از این نمایش قابل استخراج می باشند عبارتند از: Spectral centroid, Band energy ratio, Spectral flux, ...

همانطور که گفته شد، در هر یک از حوزه های بالا، تنها ویژگی های مربوط به همان حوزه قابل استخراج است. اما دسته ی دیگری از ویژگی ها وجود دارند که اطلاعاتی را در هر دو این حوزه ها در اختیار ما قرار می دهند.

۳. ویژگی های حوزه ی زمان-فرکانس^۴: برای استخراج اینگونه ویژگی ها می توان از نمایش های حوزه ی زمان-فرکانس استفاده نمود: Mel-Spectrogram, Constatn-Q transform, Spectrogram, ...

«طیف سنج ها» (Spectrograms) روش های مفیدی برای بصری سازی طیف فرکانس های یک صدا و چگونگی تغییر آن ها در طول یک بازه زمانی هستند. اسپکتروگرام از روی خروجی تبدیل فوریه زمان کوتاه بدست می آید. برای بدست آوردن اسپکتروگرام در ابتدا تبدیل فوریه زمان کوتاه روی سیگنال اعمال می شود، (سیگنال به چندین بازه زمانی تبدیل شده و روی هر کدام یک تبدیل فوریه جدا اعمال می شود) و در نتیجه آن به ازای هر بازه زمانی یک طیف فرکانسی بدست می آید. سپس ضرایب هر کدام از طیف های فرکانسی براساس میزان دامنه ای که دارند به یک کد رنگی تبدیل می شوند. که در نتیجه آن ما یک نقشه رنگی برحسب زمان و فرکانس بدست می آید که ما می توانیم با تحلیل آن متوجه شویم که طیف فرکانسی سیگنال در طول زمان به چه صورت تغییر می کند.

¹ Time domain features

² Waveform

³ Frequency domain features

⁴ Time-frequency features

یادگیری ماشین

تمرین «۵»



ددلاین: ساعت ۲۳:۵۹ | ۱۴۰۲/۱۰/۱۸

دستیاران آموزشی

کامیار رحمانی

بهناز ریوندی

امیرمحمد کویش پور

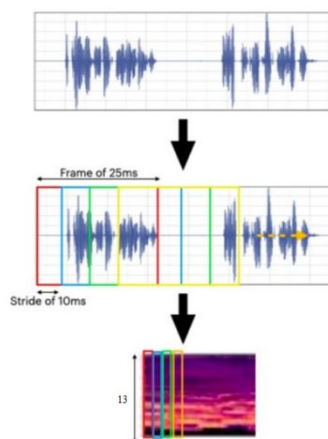
دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

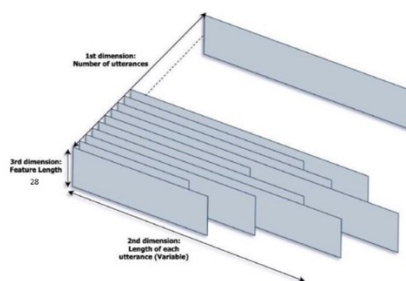
نیم‌سال اول ۱۴۰۳-۱۴۰۲

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 $(T \times 28)$ 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



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: راه‌حلی که می‌توانیم استفاده کنیم این است که برداری به نام «بردار زمینه» به اندازه K در اطراف هر بردار اضافه کنیم.)

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

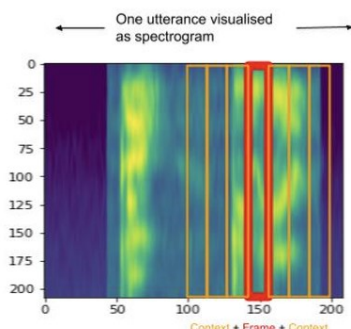
نیم سال اول ۱۴۰۳-۱۴۰۲

برای مثال، بردار زمینه با اندازه ۵ به این معنی می باشد که ما یک ورودی با اندازه (۲۸، ۱۱) به شبکه ارائه می دهیم - اندازه ۱۱ ممکن است به صورت زیر توضیح داده شود: بردار برای پیش بینی برچسب از اطلاعات ۵ بردار قبل از این بردار، و ۵ بردار بعد از آن می تواند استفاده کند.

راه های مختلفی برای پیاده سازی 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، می خواهیم قبل و بعد از این نمونه، padding صفر را اضافه کنیم تا به بعد (۲۸، ۱۰۱۰) تبدیل شود. به همین ترتیب لایه ورودی (۲+۱*اندازه زمینه)*۲۸ گره خواهد داشت. (چرا؟)

بردار زمینه یک فرامتر است و مقدار پیشنهادی بردار زمینه برای تنظیم این تمرین بین ۵۰-۵۰۰ است. برای اطلاعات بیشتر در مورد بردار زمینه، به پیوست مراجعه کنید



شکل ۶ حرکت طیف از فریم های مجاور

۶. تنظیم پارامترهای مدل

در یادگیری عمیق، هایپرپارامترها شامل متغیرهایی هستند که برای تنظیم شبکه عصبی استفاده می شوند، 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
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: ۰.۰۰۳، بهینه ساز: Adam

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

نیم‌سال اول ۱۴۰۳-۱۴۰۲

۵. (۱۰٪) [نظری-خوشه‌بندی] به سوال‌های پرسیده شده، به طور کامل و در صورت نیاز به تعداد گام خواسته شده الگوریتم‌ها را اجرا کنید.
۱. انتظار دارید در الگوریتم DBSCAN افزایش یا کاهش اندازه شعاع همسایگی ϵ (بدون تغییر مقدار MinPts) چگونه بر تعداد خوشه‌ها و تعداد نمونه‌هایی که نویز تشخیص داده می‌شوند، اثر بگذارد؟ چرا؟
۲. نمونه‌های زیر که توسط یک ویژگی X توصیف شده‌اند را به روش سلسله‌مراتی پایین به بالا (با استفاده از معیار فاصله Single Linkage) به دو خوشه تقسیم کنید و کام‌های خوشه‌بندی 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-خوشه‌بندی K-means]
۱. ثابت کنید الگوریتم k-means همگرا می‌شود.
۲. برای دادگانی با n داده و k خوشه ($k > 2$)، نصف داده‌ها در ناحیه متمرکز و نصف دیگر در ناحیه‌ای با چگالی کمتر قرار گرفته‌اند و این دو ناحیه تقریباً جدا از هم‌اند، پس از خوشه‌بندی دادگان (با احتساب MSE) آیا در نهایت مرکز خوشه‌ها به صورت یکنواخت بین دو ناحیه ذکر شده توزیع می‌شوند؟ یا در ناحیه‌ای 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 نقطه‌ای با بیشترین احتمال به دست آمده از فرمول زیر باشد. مراکز بعدی نیز به همین صورت. در انتخاب مرکز در مراحل بعدی (هنگام به‌روز کردن مراکزها) نیز نقطه با بیشترین احتمال فرمول زیر را به عنوان مرکز جدید بر می‌گزینیم.
- $$\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$$
- $D(x)$ طول کوتاهترین فاصله بین داده x تا نزدیک‌ترین مرکز است که قبلاً انتخاب شده است. X نیز مجموعه تمام داده‌ها است. نتیجه خوشه‌بندی این الگوریتم را نسبت به نتیجه k-means مقایسه کنید؟ از لحاظ سرعت همگرایی نیز مقایسه کنید؟
۴. آیا استفاده از معیار فاصله کوسینوسی در k-means برای داده‌های شامل ویژگی زمان (Time Series Data) معیار مناسبی است؟ توضیح دهید؟ و اگر مناسب نیست، چه معیار اندازه‌گیری فاصله‌ای مناسب‌تر است؟