

# تمرین چهارم درس داده‌کاوی

خوشه‌بندی و شبکه عصبی

زمستان ۹۸

## ۱ تمرین‌های تئوری

سوال ۱

شرط همگرایی الگوریتم یادگیری پرسپترون چیست؟

سوال ۲

برای هر تابع منطقی زیر، یک مدل پرسپترونی ارائه کنید که بتواند آن عملیات را پیاده کند. در هر مورد، وزن‌ها و بایاس‌ها را تعیین کرده و مرز تصمیم‌گیری را نمایش دهید.

NOR •

NAND •

XNOR •

سوال ۳

الگوریتم `kmeans` بخاطر سادگی و قدرت قابل قبول خود مورد توجه محققان بوده است و برای آن `variation` های مختلفی (مانند `kmeans++`) ارائه شده است. یکی از این `variation` ها را به دلخواه انتخاب کنید و در حد یک پاراگراف توضیح دهید که تفاوت آن با نسخه‌ی اصلی چیست و باعث چه بهبودهایی می‌شود.

سوال ۴

چشم انسان، در فضای ۲ بعدی بطور سریع و موثر کیفیت الگوریتم‌های خوشه‌بندی را تشخیص می‌دهد. آیا می‌توانید روشی برای بصری سازی اطلاعات در ۳ بعد ارائه دهید که به انسان برای تشخیص کیفیت خوشه‌ها کمک کند؟ برای ابعاد بالاتر چگونه؟

## ۲ تمرین‌های عملی

### سوال ۱

۱. پیاده‌سازی الگوریتم پرسپترون:

برای آشنایی بیشتر با الگوریتم یادگیری پرسپترون به اینجا مراجعه کنید. هدف نهایی مدلسازی در این مساله، یافتن بردار وزن‌هاست.

---

**Algorithm:** Perceptron Learning Algorithm

---

```
P ← inputs with label 1;
N ← inputs with label 0;
Initialize w randomly;
while !convergence do
    Pick random  $\mathbf{x} \in P \cup N$ ;
    if  $\mathbf{x} \in P$  and  $\sum_{i=0}^n w_i * x_i < 0$  then
        |  $\mathbf{w} = \mathbf{w} + \mathbf{x}$ ;
    end
    if  $\mathbf{x} \in N$  and  $\sum_{i=0}^n w_i * x_i \geq 0$  then
        |  $\mathbf{w} = \mathbf{w} - \mathbf{x}$ ;
    end
end
//the algorithm converges when all the inputs are
classified correctly
```

---

راهنمای قدم به قدم برای مدلسازی این الگوریتم در فایل Tutorial 1 همراه تمرین آپلود شده است که توصیه می‌شود از آن استفاده نمایید. برای دسترسی به دیتاست این تمرین، اینجا را ملاحظه کنید.

۲. طراحی پرسپترون چند لایه:

در ادامه‌ی بخش قبلی که در آن سازوکار پرسپترون را طراحی کرده‌اید، یک شبکه عصبی با ۱ لایه پنهان (یک لایه ورودی، یک لایه پنهان و یک لایه خروجی) را طراحی کنید و کارکرد آن روی جداسازی داده‌ها و دقت آن را با پرسپترون مقایسه کنید. نتایج قسمت اول و دوم را کامل در گزارش خود بیاورید. برای آشنایی با ریاضیات این مسئله این لینک را مطالعه کنید. در حل این مسئله می‌توانید برای مشتق‌گیری از gradient tape کمک بگیرید ولی لایه‌ها را باید خودتان پیاده سازی کنید و از لایه آماده dense استفاده نکنید. (مشابه کاری که در نوتبوک Session7 انجام شد.)

### سوال ۲

در این مسئله به مقایسه روش‌های معروف خوشه‌بندی می‌پردازید و برای این کار ۲ دیتاست به همراه تمرین در اختیار شما قرار گرفته است.

۱. ابتدا این دو مجموعه داده را با استفاده از الگوریتم‌های kmeans و dbscan خوشه‌بندی کنید. تعداد مناسب خوشه‌ها برای kmeans را با روش elbow به دست بیاورید. همچنین برای روش dbscan پارامترها را طوری تنظیم کنید که علاوه بر خوشه‌بندی، داده‌های پرت نیز تشخیص داده‌شوند و نتیجه را رسم کنید. اگر الگوریتمی در خوشه‌بندی مناسب یک دیتاست موفق نبود علت آنرا توضیح دهید.

۲. با رسم نمودار dendrogram تعداد خوشه‌های مناسب را برای هر دیتاست پیدا کنید و با استفاده از یکی از روش‌های سلسله‌مراتبی (agglomerative یا divisive) خوشه‌بندی را انجام دهید. مزایا و معایب روش سلسله‌مراتبی را بیان کنید. استفاده از کتابخانه scikit learn کمک بسیاری به شما خواهد کرد.

### سوال ۳

در این مسئله به دسته‌بندی داده‌های MNIST با استفاده از شبکه‌های عصبی می‌پردازید. برای به دست آوردن دیتاست می‌توانید از این لینک استفاده کنید.

۱. در ابتدا این داده‌ها را با استفاده از شبکه fully connected یا همان dense در keras یا tensorflow دسته‌بندی کنید. سعی کنید تعداد لایه و نوروهای لازم در هر لایه را برای به دست آمدن بالاترین دقت در داده‌های تست پیدا کنید.

۲. در این قسمت همان طبقه‌بندی را با استفاده از یک شبکه کانولوشنی انجام دهید. به عنوان معماری می‌توانید از مثال‌های کلاس تدریس‌یاری مانند AlexNet استفاده کنید ولی باید کد آنرا بنویسید و مستقیماً وزن‌های مدل آموزش دیده را load نکنید. دقت به دست آمده در روش اول و دوم را با هم مقایسه کنید و علت آن را تحلیل کنید.

### توجه:

- از کامنت‌های مناسب برای بیان بخش‌های مختلف کدتان استفاده کنید.
- در فایل بارگذاری شده بهترین نتیجه خود را گزارش و کد مربوط به تولید همین نتیجه را قرار دهید.
- در مورد مدل‌ها، ویژگی‌ها و پارامترهای استفاده شده توضیحات و تحلیل‌های مد نظرتان را گزارش کنید.
- رعایت کردن تمامی مراحل پیش پردازش داده‌ها، مهندسی فیچرها، انتخاب بهترین فیچر، ساخت مدل‌های مختلف بر اساس تناسب آن‌ها با داده‌ی موردنظر و مقایسه‌ی آن‌ها و مراحل مورد نیاز دیگر، الزامی است و از آنجا که راه‌حل‌های مختلفی از این مسائل در اینترنت موجود است، نمره‌ی اصلی این تمرین مربوط به گزارش شما از روند حل مساله‌تان خواهد بود.
- پاسخ بخش تئوری و گزارش بخش عملی را در قالب pdf به نام "ID\_CNN.pdf" به همراه فایل‌های کد در قالب یک فایل zip. به نام "ID\_CNN.zip" آپلود کنید. مثلاً اگر شماره دانشجویی شما ۹۶۱۳۱۹۰ می‌باشد، فایل را 9613190\_CNN نام‌گذاری کنید.
- در صورت وجود سوال یا ابهام با ایمیل‌های yasaman.m.1997@gmail.com و heidarymm@yahoo.com در ارتباط باشید.