



دانشگاه صنعتی امیرکبیر

(پلیتکنیک تهران)

دانشکده ریاضی و علوم کامپیوتر

تمرین چهارم درس داده کاوی محاسباتی

عنوان تمرین: پیش بینی اعداد دست نویس به کمک حل کمترین مربعات

استاد درس :

دکتر فاطمه شاکری

آبان 1402

بخش اول: سوال تشریحی

پیچیدگی محاسباتی، محاسبه تجزیه QR با استفاده از ماتریس های هاوس-هولدر را به دست آورید.

نکات:

1. برای انجام تمرین بالا نیازی به ارسال پاسخ به صورت تایپ شده نمی باشد. ارسال تصویر پاسخ های دست نوشته خود نیز کافی است.

بخش دوم: پیاده سازی

فرض کنید مجموعه ای از ارقام دست نوشته به عنوان مجموعه ی آموزشی ذخیره شده اند. هر رقم یک تصویر سیاه و سفید است که متناظر با آن می توان یک ماتریس $m \times n$ در نظر گرفت. هر درایه از این ماتریس متناظر با یک پیکسل از تصویر می باشد که با توجه به رنگ پیکسل مورد نظر، یک عدد از 0 تا 255 در این درایه قرار می گیرد. حال اگر ستون های این ماتریس را به ترتیب زیر هم قرار دهیم، یک بردار mn تایی حاصل می شود. بنابراین متناظر با هر تصویر یک بردار mn تایی داریم. حال قصد داریم برای یک رقم دست نوشته ی جدید، مشخص کنیم متعلق به کلاس مربوط به کدام عدد می باشد. Z را بردار مربوط به رقم دست نوشته ی جدید در نظر می گیریم.

ابتدا تعدادی از ارقام 0 موجود در مجموعه داده های آموزشی را در یک ماتریس به نام A_0 ذخیره می کنیم، به عبارت دیگر هر ستون در ماتریس A_0 ، متناظر با بردار تصویر مربوط به یکی از 0 های ذخیره شده در مجموعه داده ی آموزشی می باشد. به همین ترتیب تعداد برابری از ارقام 1 موجود در مجموعه ی آموزشی را در ماتریس A_1 و به همین ترتیب ماتریس های متناظر با ارقام دیگر را ذخیره می کنیم. حال برای هر ماتریس A_i ، فاصله ی بردار Z را تا فضای ستونی ماتریس A_i را به دست آورده و آن را α_i می نامیم:

$$\alpha_i = \min (\|Z - Ax\|_2), \quad i = 0, \dots, 9$$

اگر $\alpha_k = \min(\alpha_i)$ ، پیش بینی می کنیم رقم دست نوشته جدید از مجموعه داده آزمایشی، رقم k می باشد.

1. برای حل مسئله کمترین مربعات بالا از تجزیه QR با استفاده از ماتریس های هاوس-هولدر که خودتان پیاده سازی کرده اید استفاده کنید (از ستورات آماده مانند np.linalg.qr یا دستورات مشابه در کتاب خانه های دیگر استفاده نشود)

2. با توجه به توضیحات بالا، داده های موجود در مجموعه دادگان آزمایشی را، در دو حالتی که از پیاده سازی خود برای محاسبه تجزیه QR استفاده شده و حالتی که تجزیه QR را به کمک دستورات آماده NumPy به دست آورده اید، پیش بینی کنید و ماتریس آشفستگی را برای پیش بینی های خود در این دو حالت بدست آورده و آن را گزارش دهید.

3. تاثیر تعداد داده های آموزشی انتخاب شده در تشکیل ماتریس های A_i را روی دقت خروجی الگوریتم بررسی کنید و در گزارش کار خود تحلیل کنید. برای این کار کافی است تمام الگوریتم به ازای ماتریس ها A_i با تعداد ستون مختلف اجرا کنید. (برای انجام این بخش می توانید از دستورات آماده استفاده کنید)

نکات:

1. از آن جا که پیاده سازی تجزیه QR با استفاده از ماتریس های هاوس-هولدر به عهده خودتان می باشد یکی از چالش موجود در این تمرین پیچیدگی محاسباتی بالا می باشد. از این رو سعی کنید تا در پیاده سازی های خود تا حد امکان از حلقه یا *for* کمتر استفاده کنید و محاسباتی خود را به صورت برداری انجام دهید. راهنمایی: یکی از حلقه هایی که حذف آن می تواند تاثیر چشمگیری در بهبود پیچیدگی پیاده سازی شما داشته باشد این است که، به جای پیش بینی داده های آزمایشی در یک حلقه به ازای هر داده آزمایشی و تبدیل هر تصویر آزمایشی به بردار Z و انجام پیش بینی برای آن تصویر، پیش بینی تمام داده های آزمایشی به صورت همزمان و به کمک تولید ماتریس Z انجام بپذیرد. در واقع هر ستون ماتریس Z متناظر با بردار یکی از تصاویر مجموعه دادگان آزمایشی می باشد.