

به نام خدا

تمرین سری سوم (موعد تحویل ۴ شنبه ۱۶ فروردین ساعت ۵ بعد از ظهر)

بخش اول

میخواهیم یک مسئله طبقه بندی دو کلاسه را با استفاده از رویکرد CSP بر روی داده های EEG انجام دهیم. کلاس اول مربوط به تصور حرکت پا و کلاس دوم مربوط به تصور انجام یک سری عملیات تفریق ذهنی است.

سه ماتریس سه بعدی و یک بردار در فایل hw3-1.mat به شرح زیر در اختیار شما قرار داده شده است:

داده های آموزش مربوط به کلاس اول در یک ماتریس سه بعدی با نام TrainData_class1 به ابعاد 256×60 (تعداد کانالها ۳۰، تعداد نمونه های زمانی ۲۵۶ و تعداد آزمایشها ۶۰) قرار داده شده است. برای همه ی آزمایش ها، میانگین هر کانال را صفر کنید و در نهایت داده های جدید را در TrainData_class1 ذخیره کنید.

داده های آموزش مربوط به کلاس دوم در یک ماتریس سه بعدی با نام TrainData_class2 به ابعاد 256×60 (تعداد کانالها ۳۰، تعداد نمونه های زمانی ۲۵۶ و تعداد آزمایشها ۶۰) قرار داده شده است. برای همه ی آزمایش ها، میانگین هر کانال را صفر کنید و در نهایت داده های جدید را در TrainData_class2 ذخیره کنید.

داده های آزمون در یک ماتریس سه بعدی با نام TestData به ابعاد $256 \times 40 \times 30$ (تعداد کانالها ۳۰، تعداد نمونه های زمانی ۲۵۶ و تعداد آزمایشها ۴۰) قرار داده شده است. برای همه ی آزمایش ها، میانگین هر کانال را صفر کنید و در نهایت داده های جدید را در TestData ذخیره کنید.

کلاس داده های آزمون در یک بردار با نام TestLabel به ابعاد 1×40 قرار داده شده است.

قسمت های الف تا ج فقط با استفاده از داده های آموزش انجام می شود و کاری با داده های آزمون ندارد

الف) با استفاده از داده های آموزش، فیلترهای مکانی CSP (W_{csp}) را به دست آورده و بر روی داده های آموزش اعمال کنید. حتما ستون های W_{csp} را نرمالیزه کنید. سیگنال فیلتر شده متناظر با فیلتر اول ($W_{csp}(:,1)$) و آخر ($W_{csp}(:,30)$) را برای آزمایش ۴۹م کلاس اول و همچنین برای آزمایش ۴۹م کلاس دوم رسم کنید (مشابه شکلی که در کلاس کشیده شد). همان طور که مشاهده می کنید برای هر آزمایش، خروجی یک فیلتر پراکندگی زیادی دارد و خروجی فیلتر دیگر پراکندگی کمی دارد. به صورت کمی هم پراکندگی ها (واریانس ها) را (فقط برای این دو آزمایش!) گزارش کنید.

ب) قدر مطلق فیلترهای مکانی اول و آخر به دست آمده در قسمت الف را رسم کنید. برای هر کلاس، کدام کانال (یا کانال ها) تاثیر بیشتری در خروجی فیلتر دارند؟ در واقع در اینجا متوجه می شویم اثر هر کدام از تصوراتی که فرد در کلاس ۱ و ۲ داشته، در کجای مغز بیشتر است.

ج) واریانس داده های فیلتر شده را با در نظر گرفتن ۱۴ فیلتر مکانی مهمتر (متناظر با فیلترهای شماره ی ۱ تا ۷ و فیلترهای شماره ۲۴ تا ۳۰) استخراج کنید. در واقع هر داده ی آموزش یک ماتریس 30×256 بوده است که در اینجا به یک بردار با ۱۴ درایه تبدیل (و یا فشرده) می شود. حال با استفاده از طبقه بند LDA بهترین تبدیل خطی (WLDA) و مرز متناظر (C) که داده های دو کلاس را از هم جدا می کند، پیدا کنید. حتما WLDA را نرمالیزه کنید. توجه داشته باشید که WLDA یک بردار با ۱۴ درایه است و C یک اسکالر است. مقدار WLDA و C به دست آمده را گزارش کنید.

د) حال به سراغ داده های آزمون می رویم. در ابتدا فرض کنید برچسب (کلاس) داده های آزمون را ندارید. با استفاده از ۱۴ فیلتر مکانی مهمتر که از داده های آموزش به دست آمده بودند و همچنین WLDA و C که آنها هم از داده های آموزش استخراج شده بوند، درباره ی برچسب داده های آزمون تصمیم گیری کنید.

ه) نتیجه به دست آمده از قسمت د را با برچسب واقعی داده های آزمون مقایسه کنید. برچسب چه تعداد از داده های آزمون را درست تخمین زدید؟ برچسب واقعی و همچنین تخمین زده شده ی داده ها را در یک شکل رسم کنید. محور X نمایانگر شماره ی داده آزمون باشد و محور Y برچسب واقعی و تخمینی آن را مشخص کند.

بخش دوم)

در این تمرین می خواهیم روش CCA را برای شناسایی فرکانس تحریک داده های SSVEP استفاده کنیم (steady state visually evoked potentials). یک ماتریس سه بعدی و دو بردار در فایل hw3-2.mat به شرح زیر در اختیار شما قرار داده شده است:

- فرکانس های تحریک مورد استفاده در آزمایش در بردار freq با ابعاد 1×5 قرار دارند.
- داده های ضبط شده در ماتریس data با ابعاد $15 \times 1250 \times 6$ (تعداد کانالها ۶، تعداد نمونه های زمانی ۱۲۵۰ و تعداد آزمایشها ۱۵) قرار دارند. فرکانس نمونه برداری ۲۵۰ هرتز بوده است و یا به عبارت دیگر طول زمانی هر آزمایش ۵ ثانیه می باشد.
- برچسب واقعی داده ها در بردار label با ابعاد 1×15 قرار دارند. در واقع با استفاده از این بردار متوجه می شویم هر داده با تحریک چه فرکانسی ضبط شده است.

با استفاده از روش CCA فرکانس تحریک هر داده را تخمین زده و جواب تخمین زده را با مقدار واقعی آن مقایسه کنید. برای ایجاد ماتریس های template هارمونیک های تا ۴۰ هرتز را در نظر بگیرید.

به نکته ی زیر توجه داشته باشید:

- در این تمرین رویکرد یادگیری ماشین برای تخمین برچسب داده ها در نظر نگرفته شده است!