

باسمه تعالی

- تمرین سری دوم درس مبانی هوش محاسباتی
- تمامی فایل های مرتبط با تمرین را در قالب یک فایل فشرده با نام HW2_studentNumber.zip در مودل درس بارگذاری نمایید.
- مهلت ارسال تمرین تا ساعت ۲۳:۵۵ روز شنبه مورخ ۲۵ آبان ۱۳۹۸ می باشد.
- در صورت وجود هرگونه سوال می توانید با ایمیل زیر در ارتباط باشید.

ci.fall2019@gmail.com

در تمرین دوم قرار است یک framework تکاملی طراحی کنید و کاربرد آن را در حل یک مسئله ببینید.

فایل های پایتون برای نمونه ضمیمه ای این تمرین شده اند. زبان برنامه نویسی آزاد است و رعایت ساختار داده شده اجباری نیست، در صورت تمایل می توانید از زبان های دیگر و هر ساختار دلخواهی استفاده کنید.

تعریف مسئله:

در این مسئله قرار است یک مسئله کاهش بعد را با روش PCA^۱ حل کنیم. کاهش بعد عملیاتی است که برای مجموعه داده های با ابعاد بزرگ، جهت بهبود الگوریتم های داده کاوی و آماری از نظر زمان اجرا و فضای مصرفی استفاده می شود.

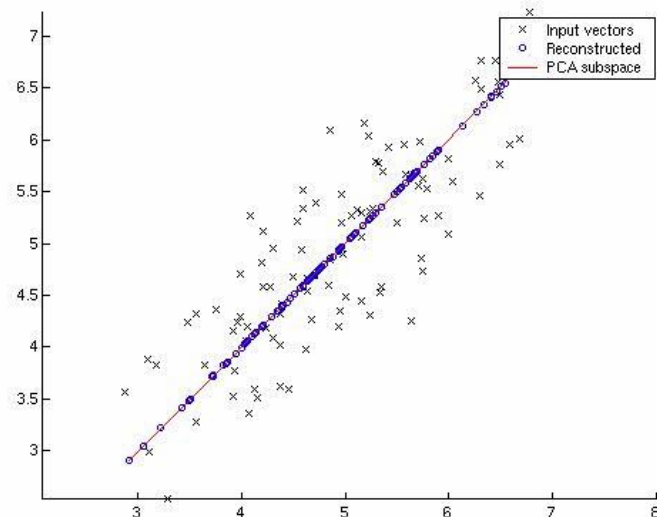
مسئله PCA با استفاده از ماتریس کواریانس قابل حل است که در صورت علاقه می توانید از [اینجا](#) مطالعه فرمایید. اما در این تمرین قصد داریم از طریق استراتژی تکاملی یک نمونه بسیار ساده از این مسئله را حل نماییم.

برای سادگی مسئله یک مجموعه داده دو بعدی در نظر میگیریم که می خواهیم آن را کاهش بعد داده و به یک مجموعه داده یک بعدی تبدیل نماییم و داده ها را روی یک بعد (یک ترکیب خطی از دو بعد) نظیر کنیم به طوریکه

¹ Principal Component Analysis

داده های نظیر شده بیشترین انحراف معیار را با هم داشته باشند تا تمایز بین داده ها بیشترین مقدار شود و بهترین کاهش بعد را داشته باشیم.

شکل ۱ را در نظر بگیرید که نمونه ای از این مسئله است:



شکل ۱

نحوه کاهش بعد

همانطور که اشاره شد برای تبدیل دو بعد به یک بعد به دنبال یافتن بهترین ترکیب خطی ممکن از این دو بعد هستیم که داده های نظیر شده روی آن بعد بیشترین انحراف معیار را با هم داشته باشند.

حال برای کاهش بعد نقطه ی (x_i, y_i) در این مجموعه داده، یک ترکیب خطی با ضرایب a و b خواهیم داشت که z_i مقدار نظیر شده روی تک بعد ایجاد شده است:

$$z_i = ax_i + by_i$$

حال برای همه ی نقاط موجود در مجموعه داده این تبدیل را حساب کرده و بین z_i های مختلف انحراف معیار را به دست می آوریم و این مقدار میتواند به عنوان یک معیار شایستگی یا عدم شایستگی تبدیل در نظر گرفته شود.

نگاهی به ساختار پیشنهادی:

کلاس Chromosome :

فیلدهای این کلاس عبارت است از تعداد ژن‌های کروموزوم، لیستی از ژن‌ها، کمترین مقدار و بیشترین مقداری که ژن‌ها میتوانند داشته باشند.

حال برای کانستراکتور این کلاس باید لیستی از ژن‌ها به تعداد ژن‌های کروموزوم ایجاد کنید که مقداری به صورت تصادفی بین min و max دارند.

تابع evaluate یک کروموزوم به عنوان ورودی می‌گیرد و بسته به نوع و شرایط مسئله داده شده، شایستگی کروموزوم را به عنوان خروجی برمیگرداند.

فایل ES:

در این فایل، قالب مجموعه‌ای از توابع پیشنهادی و هایپر پارامترهای تنظیم کننده برای پیاده سازی استراتژی تکاملی ارائه شده است.

فایل plot:

این فایل شامل توابعی است که برای رسم داده‌های دیتاست و بردار به دست آمده توسط خروجی الگوریتم تکاملی کاربرد دارند.

فایل file_handler:

شامل توابعی برای خواندن از دیتاست و ذخیره تصویر خروجی است.

نکات تکمیلی:

۱. برای این تمرین یک گزارش مورد نیاز است که شامل موارد زیر باشد:
 - نحوه بازنمایی مسئله
 - نحوه ترکیب کروموزم‌ها با همدیگر و مقدار احتمال ترکیب و تاثیر آن در سرعت همگرایی.
 - نحوه جهش و نحوه تعیین گام جهش و تاثیر جهش در سرعت همگرایی.
 - نحوه انتخاب والدین و بازماندگان.
 - تعداد جمعیت و تعداد فرزندان (مقادیر مختلف را امتحان کرده و سرعت همگرایی را مشاهده کرده و نتیجه را ارائه دهید)
 - شرط خاتمه الگوریتم تکاملی.
۲. لازم است در طول اجرای الگوریتم، در هر نسل مقدار بهترین و بدترین شایستگی موجود در نسل و میانگین شایستگی‌های جمعیت چاپ شود.
۳. پس از اتمام اجرای الگوریتم، لازم است نقاط دیتاست، بردار به دست آمده و نقاط تصویر شده روی بردار دقیقاً مشابه شکل ۱ رسم و ذخیره شود.