

به نام خدا

علی عباس حقیقت

220797069

گزارش کار تمرین دوم درس داده
کاوی :

SStream Clustering

عرض سلام و احترام دارم خدمت شما :

هدف از بنده در اینجا نوشتن یک گزارش از عملکرد خود در انجام تمرین دوم درس داده کاوی است و توضیحات مبسوط بسیار بسیار کامل در متن مقاله و اگر برای فهم آسانتر در کامنت هایی که بنده در خود پروژه به ازای تک تک توابع و کلاس ها نوشته ام موجود است . لذا در اینجا مراحل انجام پروژه و همچنین چالش ها و مشکلاتی که با آنها روبرو شدم را به اختصار بیان خواهم کرد .

در مرحله اول انجام پروژه بنده به مطالعه دقیق مقاله پرداختم که باعث شد به درک متوسطی از نحوه خوشه بندی جریان داده برسم ، اما در گام بعد برای پیاده سازی به مطالعه عمیق و تطبیقی شبه کد موجود در مقاله با توضیحات ارائه شده در مقاله پرداختم که موجب درک بهتر خود کد و مفهوم SOStream clustering برای بنده شد .

سپس برای پیاده سازی کد و به عنوان ستون فقرات کد از پیاده سازی SOStream Clustering موجود در گیتهاب بهره بردم که تا حد بسیار زیادی مسیر رسیدن از شبه کد (موجود در مقاله) به کد را برای بنده سهل کرد و در حکم میانبر بود .

در گام بعدی کد را با توجه به اهداف خود از جمله موارد خواسته شده در متن تمرین و کلاستر های جدید به روز رسانی کردم (از جمله افزودن کلاس main برای بهره بردن از تمام کلاس و توابع و ...) و تغییرات مورد نیاز را انجام دادم.

سپس در گام بعدی تمام بخش های کد را با توجه به فهم خود کامنت گذاری کردم.

اما در کد بنده با چالشهایی هم مواجه شدم . اولین موردی که خودنمایی میکرد بخش کار با کلاس های متعدد و شی گرایی موجود در ذات الگوریتم بود که بسیار سریع با منطق پشت آن خو گرفتم .

مرحله بعدی کار با خود دیتاست ها و خوراندن آنها به کد بود که آنها هم مشکل زیادی ایجاد نکرد.

اما مهمترین مشکلی که در گام بعد ایجاد شد بخش رسم نقاط بود که در حالت دو بعدی با مشکلات زیادی همراه نبود به غیر از اینکه چاپ تدریجی در حلقه ای که جریان داده را دریافت میکرد به درستی انجام نمیشد اما عملکرد کلی آن خوب بود .

در مرحله بعدی چون داده ما سه بعدی بود و scatter3D هم لازم داشت مشکلی ایجاد شد و آنهم این بود که وقتی یک پوینت

را چاپ میکرد در همان فیگور وقتی یک داده دیگر از جریان داده خوانده میشد و حال میخواست آنرا هم چاپ کند پوینت قبلی را پاک میکرد لذا در پلات پایانی تصویر تجمعی از داده ایجاد نمیشد اما مراحل را به درستی نمایش میداد .

برای حل موارد بالا بنده راه حل های نت را امتحان کردم ولی هیچ کدام راه حل کاملی ارائه نمیداد .

یک راه حل هم در نت `plt.show` کردن در خارج از حلقه داده ها بود که تمام پلات ها را باهم به نمایش بگذارد که به لحاظ منابع رم (البته احتمالا) به مشکل برمیخورد و بدون ارور خاصی برنامه و سیستم همزمان کرش میکردند.

سپاس از زحمات بی دریغ شما

