

فاز سوم پروژه هوش محاسباتی

پاییز 1402

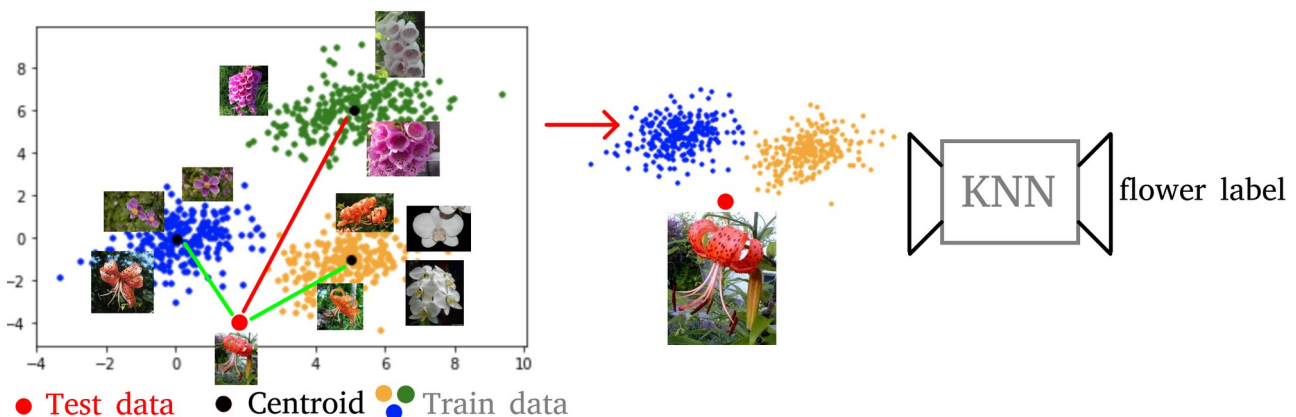
مقدمه

در این پروژه قصد داریم با استفاده از ویژگی‌های (features) تهیه شده از تصاویر دیتاست flower102، ابتدا به صورت بدون‌ناظر (unsupervised) با الگوریتم k-mean دیتاها را به صورت مناسب طبقه بندی کرده و سپس در بخش بعد با شکستن داده ها به گروه های کوچکتر و الگوریتم K-Nearest Neighbors بهترین امتیاز (score) را بدست آوریم.

در ابتدا شما باید دیتاست موجود در درایو را با استفاده از کد قرار داده شده لود کنید. این دیتاست شامل ویژگی‌هایی است که از تصاویر دیتاست flower102 بدست آمده و در نتیجه ۱۰۲ کلاس مختلف خواهید داشت و همچنین هر داده با یک بردار ۵۱۲ بُعدی ویژگی نمایش داده می‌شود. این داده‌ها شامل دو بخش آموزش و تست می باشند که 4094 داده برای آموزش و 4095 داده برای تست در نظر گرفته شده است.

پیاده‌سازی

قصد داریم از الگوریتم K-Nearest Neighbors (KNN) استفاده کرده و یک مسئله کلسیفیکیشن برای گل‌ها تعریف کنیم اما زمان مورد نیاز برای تشخیص (prediction) الگوریتم برای ما اهمیت دارد. با توجه به تعداد زیاد داده و ابعاد هر داده تصمیم داریم که دیتاست خود را با استفاده از الگوریتم k-means به چند دیتاست کوچکتر بشکنیم. برای اینکار شما باید دیتای آموزش خود را با استفاده از الگوریتم k-means به 50 (k=50) کلاستر تقسیم کنید. سپس مراکز (centroid) این کلاسترها را بدست آورده و همچنین دیتای موجود در هر کلاستر را جدا کنید. به این ترتیب شما ۵۰ دیتاست کوچکتر دارید که متناظر با هر کدام یک مرکز کلاستر نیز وجود دارد.



درنهایت برای هر داده تست باید n تا مرکز نزدیکتر را انتخاب کرده و با استفاده از الگوریتم KNN و n تا دیتاست بدست آمده، لیبل آن دیتا را بدست آورید. دقت کنید که به ازای هر یک کلاستر اضافه که استفاده می‌کنید، پناستی 0.2 - دریافت می‌کنید که از امتیاز شما کم می‌شود. معیار شما باید بیشینه کردن مقدار امتیاز باشد، این امتیاز به صورت پایین محاسبه می‌شود:

$$p(\text{penalty}) = -0.2$$

$$\text{score} = \text{accuracy} + (n * p)$$

که مقدار accuracy یک عدد اعشاری بین 0 و 100 می‌باشد. بهترین امتیاز را محاسبه کرده و دلیل آن را در گزارش خود استدلال کنید. (استدلال شما باید توجیه کند که چرا n بزرگ/کوچک باعث امتیاز بهتر/بدتر شده است و در چه شرایطی این اتفاق نمی‌افتد)

بعد از رسیدن به بهترین امتیاز، بررسی کنید که آیا ارتباطی بین لیبل‌های KNN و لیبل‌های کلاسترها وجود دارد یا خیر؟ به عبارت دیگر باید بررسی کنید که هر کلاستر چه مقدار در پیشبینی الگوریتم KNN کمک کننده بوده و توانسته کلاستری با خلوص بیشتر را بدست آورد.

همچنین برای ارزیابی K-mean می‌توانید از معیارهای ارزیابی که در درس فراگرفته اید استفاده و پیاده‌سازی کنید.