



مبانی یادگیری ماشین - تکلیف سری چهارم

مدرس: دکتر ریاحی

پاییز ۱۴۰۲

مهلت: ۱ دی ساعت ۲۳:۵۹

مسائل تحلیلی

۱. ۲ روش برای تحلیل خروجی الگوریتم های Unsupervised در زمانی که برچسب دیتا را نداریم ارائه دهید.
۲. یکی از روش های کاهش تعداد رنگ های درون یک تصویر، اعمال الگوریتم K-Means بر روی مقدار عددی هر کدام از پیکسل ها است. توضیح دهید استفاده از الگوریتم DBSCAN به جای K-Means برای این منظور چه مزایا و معایبی دارد؟ از هر کدام ۱ مورد برای کفایت.
۳. الگوریتم PCA یکی از مهمترین ترین الگوریتم ها برای کاهش ابعاد (Dimensional Reduction) است. نحوه کارکرد این الگوریتم را به طور کامل شرح دهید. پیشنهاد می شود برای آشنایی بیشتر با این الگوریتم و نحوه کار با آن، یکی از مثال های این الگوریتم را (دمو الگوریتم K-Means با استفاده از روش PCA) به منظور کاهش ابعاد مطالعه کنید.
۴. نحوه پیاده سازی درخت تصمیم بر روی این دیتا را توضیح دهید. (نحوه انتخاب ویژگی ها و زمان توقف تقسیم داده ها را بیان کنید)

مسائل کدی

۱. در این سوال باید الگوریتم K-Means را از ابتدا پیاده سازی کرده و سپس با استفاده از این الگوریتم، تعداد رنگ های عکس را کاهش دهید. این عملیات را با تعداد رنگ های مختلف تکرار کرده و علاوه بر نمایش تصاویر حاصل، نمودار elbow خروجی را به ازای تعداد رنگ های مختلف رسم کنید. با توجه به این نمودار تحلیل کنید چه تعداد رنگی برای این تصویر مناسب تر است. همچنین به نکات زیر دقت داشته باشید:
- برای افزایش سرعت الگوریتم می توانید با استفاده از توابع opencv یا دیگر کتابخانه ها ابعاد تصویر ورودی را کم کنید (مثلا 200×200).
- نباید تصویر داده شده، GrayScale شود، بلکه باید مقدار هر کانال رنگی به عنوان یک ویژگی مجزا برای هر پیکسل در نظر گرفته شود.

- استفاده از توابع موجود در کتابخانه ای که این الگوریتم را به طور کامل پیاده سازی کرده باشد ممنوع است.
- بهتر است برای افزایش سرعت الگوریتم خود تا جای ممکن از حلقه استفاده نکرده و برای عملیات های مختلف بر روی ماتریس دیتا از توابع کتابخانه numpy استفاده کنید.



نمونه کاهش یافته عکس اصلی به عکسی تنها با ۵ رنگ مختلف.

۲. الگوریتم KNN را از ابتدا پیاده سازی کرده و بر روی این [دیتاست](#) اعمال نمایید. مقادیر مختلف k را امتحان کرده و نتایج را با یکدیگر مقایسه کنید.

۳. الگوریتم های SVM و Decision Tree را نیز بر روی دیتاست سوال قبل اعمال کرده و نتایج را باهم مقایسه کنید. برای این سوال می توانید از الگوریتم های کتابخانه scikit-learn استفاده نمایید.

۴. **(امتیازی)**. در این سوال ۲ دیتاست در اختیار شما قرار گرفته است. ابتدا دیتا را در ۲ بعد نمایش دهید، سپس با استفاده از الگوریتم DBSCAN پیاده سازی شده در کتابخانه Sklearn داده ها را خوشه بندی کنید. خوشه های تشخیص داده شده را به همراه مقادیری که الگوریتم شما Outlier برچسب گذاری کرده نمایش دهید. دیتای مورد نیاز:

[Data1](#), [Data2](#)

نکات تمرین

- در صورت هرگونه **تقلب** نمره **صفر** برای شما لحاظ می گردد.
- استفاده از زبان غیر پایتون مجاز **نیست**.

پیروز و سربلند باشید