

بسمه تعالی



گزارش کار تمرین چهارم بینایی کامپیوتر

تمرین چهارم

دانشگاه صنعتی شریف

بهار ۱۴۰۱

سوال ۱) Histogram of Oriented Gradients

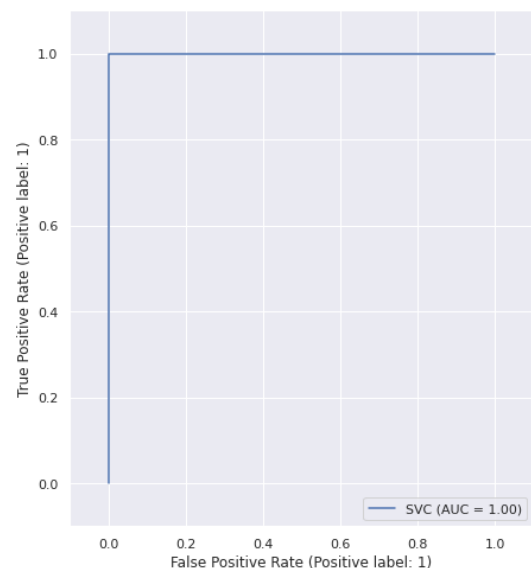
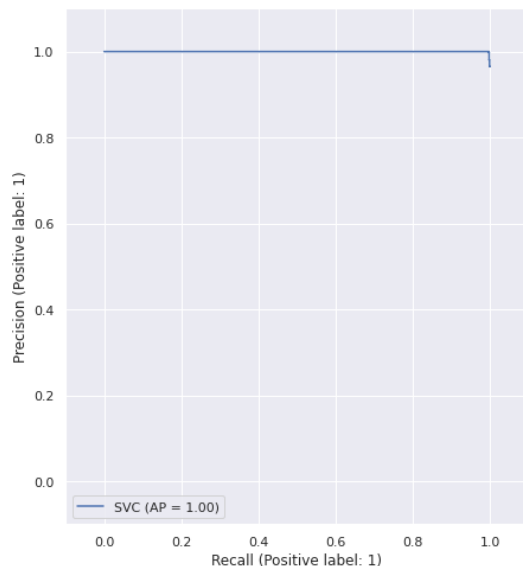
در این سوال باید با استفاده از روش **HoG** و دیتاست قید شده در گزارش، عمل تشخیص چهره را بر روی تصاویر انجام دهیم. از آنجا که برای **classifier** به دو نوع داده نیاز داریم، باید از یک دیتاست دیگر که در آن چهره وجود ندارد به عنوان داده منفی استفاده کنیم. طبق خواسته گزارش ۱۰۰۰۰ تصویر را به عنوان داده آموزش، ۱۰۰۰ داده را به عنوان داده اعتبارسنجی و ۱۰۰۰ داده را به عنوان داده تست از هر دسته جدا می‌کنیم. حال باید بردار ویژگی‌های هر یک را که از روش **HoG** به دست آمده است، استخراج کنیم. بدین منظور همه تصاویر به رزولوشن ۶۴ در ۶۴ پیکسل ریسایز شده‌اند. همچنین از تصاویر دیتاست چهره‌ها، از هر طرف ۶۲ پیکسل جدا شده تا تمرکز بیشتری روی صورت افراد قرار گیرد. اندازه هر سلول ۴ پیکسل در ۴ پیکسل است و بلوک‌ها ۲ در ۲ هستند. همچنین ۹ راستا در مجموع برای گرادیان‌ها در نظر گرفته شده است. پس در کل ۲۵۶ سلول داریم و بردار ویژگی هر تصویر ۸۱۰۰ بعد دارد.

حال با استفاده از **SVM** با استفاده از یک ابر صفحه این داده‌های مثبت و منفی را از هم جدا کنیم. برای اینکار از پارامترهای مختلف گاما و **c** و کرنل‌های متفاوت استفاده کردم و در نهایت پارامترهای زیر بهترین نتیجه را دادند.

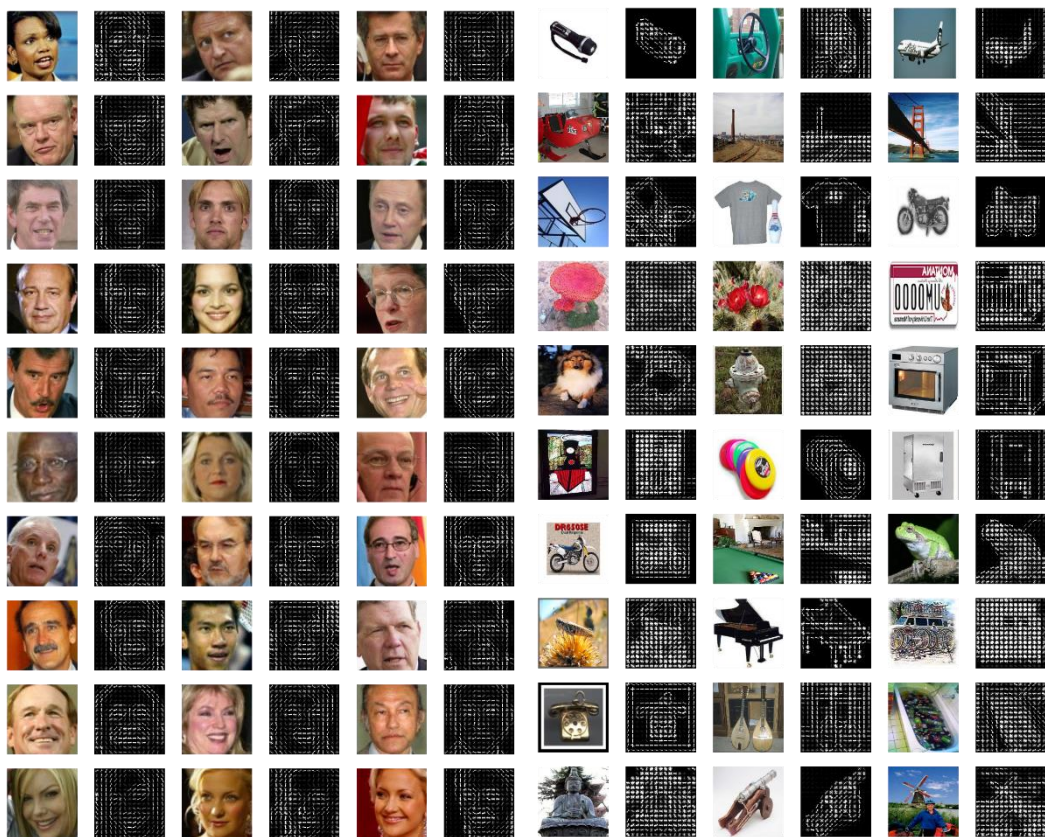
$$c = 1, \quad \text{gamma} = \text{scale}, \quad \text{kernel} = \text{rbf}$$

برای داده منفی از دیتاست **Caltech 256** به جز کلاس‌های **People** و **Faces** استفاده شده است. در نهایت پس از ترین کردن مدل **SVM**، **Accuracy** آن برای داده اعتبارسنجی برابر مقدار ۰.۹۹۵۵ شد.

امتیاز **roc_auc** و **average precision** به ترتیب برابر ۰.۹۹۹۹۴۲۹۹۹۹۹۹۹۹۹ و ۰.۹۹۹۹۴۴۶۲۲۸۷۸۸۷۲۲ شد. نمودارهای **roc** و **precision recall** هم به شکل زیر است.



چند نمونه از فیچرهای استخراج شده برای داده های مثبت و منفی به شکل زیر است.



برای تشخیص صورتها در تصویر از روش **Sliding Window** استفاده می کنیم و در هر مرحله ۱۰ پیکسل حرکت می کنیم. برای چک کردن تصویر در چند اندازه مختلف، پچهای با سایز متغیر در نظر میگیریم. برای نتیجه مطلوب در این ۳ عکس، ۵ اندازه به صورت یونیفورم از بازه ۸۰ تا ۱۸۰ انتخاب می کنیم و پچها را به شکل مربعی در آن سایز در نظر می گیریم. سپس آنها را به سایز ۶۴ در ۶۴ ریسایز کرده و با استخراج بردار ویژگی آن با همان مشخصه های گفته شده در الگوریتم **Hog**، امتیاز هر پچ را با تابع **decision_function** به دست می آوریم. ترشهولد مناسب برای این نمونهها، امتیاز بالاتر از ۱.۵۷ بود. این امتیاز نشان دهنده **confidence** مدل از پیش بینی خود است. حال ريجن های متعددی با اشتراکات نسبتا بالایی شناخته می شود، با استفاده از الگوریتم **non maximum suppression** و الویت دادن به پچهای با امتیاز بالاتر می توانیم بهترین ريجن را از بین ريجن های نشان دهنده یک صورت انتخاب کنیم. در نهایت ريجن ها را بر روی تصویر کشیده و با نامهای گفته شده ذخیر می کنیم. لازم به ذکر است، مدل به داده سلکت شده برای نمونه های مثبت و منفی که در هر بار اجرا به صورت تصادفی انتخاب می شود، تا حد کمی وابسته است و ممکن است نتایج در اجراهای متفاوت تا متغیر باشند.

