



تمرین اول یادگیری ماشین

توجه:

- کد باید فقط در زبان پایتون باشد.
- استفاده از کتابخانه‌های از پیش آماده مجاز است، مگر اینکه نقیض آن گفته شود. در گزارش کتابخانه‌ها ذکر شوند.
- لطفاً علاوه بر ضمیمه کردن کد و تصاویر، نتایج را تحلیل و در فایل گزارش خود ضمیمه کنید.
- فایل گزارش به فرمت pdf و به زبان فارسی باشد.

تمرین اول را به assignments.ml.fall2021@gmail.com ایمیل کنید.

سوال اول

با توجه به تعریف یادگیری ماشین از دیدگاه Tom Mitchell مسائل زیر را به یادگیری با نظارت و بدون نظارت دسته بندی کنید، و اگر یادگیری با نظارت است، زیرشاخه آن را ذکر کنید. سپس معیار ارزیابی و تجربه را مشخص کنید (به صورت انتزاعی کافی است، نیازی به معادلات ریاضیاتی نیست).

(آ) تشخیص نژاد اسب در تصویر

(ب) دسته بندی بازدیدکنندگان سایت برای معرفی محصولات

(ج) بررسی درخواست های به سرور و جلوگیری از درخواست های مشکوک

(د) پیش بینی میزان بارش یک منطقه در ۱۰ سال آینده

سوال دوم

در این سوال قصد داریم که با استفاده از الگوریتم پرسپترون یک مسئله دسته بندی دودویی را حل کنیم. دادگان این تمرین ضمیمه شده اند. هر نمونه از این دادگان شامل دو ویژگی x_1 و x_2 است، و ستون labels برچسب ها است. فایل train.csv دادگان آموزشی است، و فایل test.csv دادگان تست است که فاقد برچسب است.

۱. ابتدا ۵۰ عدد از دادگان را به صورت تصادفی به عنوان دادگان اعتبار سنجی انتخاب کنید.

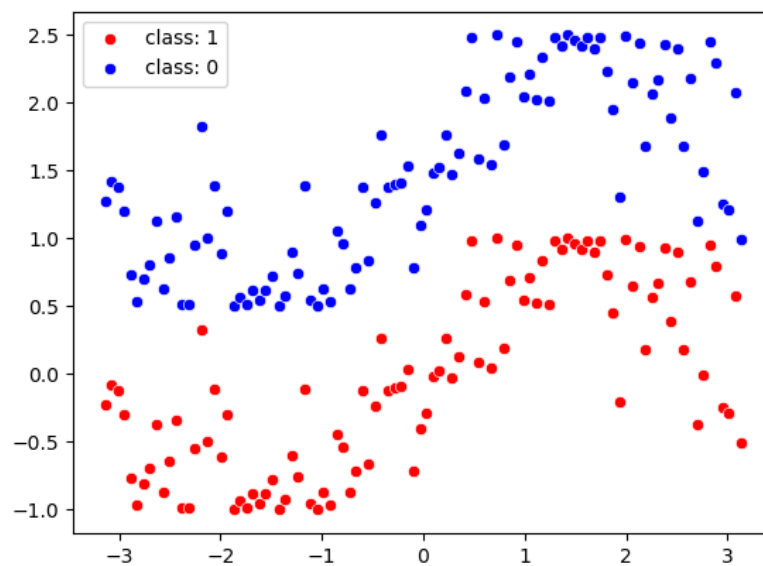
۲. دادگان را با روش مد نظرتان نرمال سازی کنید. معادلات مربوط به روش را گزارش دهید و توضیح دهید که چرا نرمال سازی می تواند تاثیر مثبت داشته باشد.

۳. اکنون دو دادگان جدید را در دو نمودار مجزا با مشخص کردن کلاس ها رسم کنید. (مشابه شکل ۱)

۴. الگوریتم پرسپترون را برای تعداد محدودی تکرار بر روی دادگان آموزشی، آموزش دهید. سپس دقت را برای دو دادگان آموزشی و اعتبارسنجی گزارش کنید. اشکال را با خط جداکننده^۱ رسم کنید (از وزن های بدست آمده استفاده کنید). در رابطه با بالا بودن و یا پایین بودن دقت صحبت کنید و دلیل آن را گزارش کنید.

۵. اکنون آزمایش بالا را با استخراج ویژگی های چند جمله ای با درجه ۲، ۳، ۵ و ۱۰ انجام دهید و در هنگام استخراج ویژگی رابطه بین متغیرها در نظر گرفته شود. با توجه به اینکه ممکن است محاسبه محور جداکننده کمی دشوار شود، کافی است در اشکال رسم شده تنها نمونه های اشتباه را نشان دهید. با توجه به نتایج بدست آمده، مدل ها را از نظر زیربرازش و بیش برازش تحلیل کنید.

۶. بهترین مدل را انتخاب کنید. از طریق آن پیشبینی را بر روی دادگان تست در دسترس انجام دهید. پیشبینی را در فایل با نام preds.csv ذخیره و ضمیمه کنید. (فایلی به نام preds.csv ضمیمه شده است که نمونه ای از فایل مورد انتظار است).



شکل ۱

¹DECISION BOUNDARY