

تکلیف شماره ۳-۲ مقدمه‌ای بر یادگیری ماشین: Naïve Bayes

تشخیص سرطان:

مجموعه داده‌های [Breast Cancer Wisconsin \(Diagnostic\) Data Set](#)، شامل مجموعه‌ای از ویژگی‌های اندازه‌گیری شده پزشکی برای تشخیص سرطان سینه می‌باشد. داده‌ها برای استفاده راحت‌تر پیش‌پردازش شده و در فایل BreastCancer.csv قرار داده شده‌اند. ابتدا داده‌ها را به صورت یکنواخت از هر دو کلاس بین داده‌های آموزش، ارزیابی و آزمون تقسیم‌بندی کنید.

الف) با استفاده از مدل Naïve Bayes و با روش Cross Validation عملکرد این الگوریتم را روی داده‌ها مشاهده نمایید و صحت Accuracy و ماتریس درهم‌ریختگی (Confusion Matrix) را گزارش کنید.

ب) این بار با استفاده از مدل Gaussian Naïve Bayes فرآیند فوق را تکرار کرده و نتایج را با حالت قبلی مقایسه کنید.

پ) امتیازی: یکی از نکات مهم در تشخیص پزشکی به کمک الگوریتم‌های هوشمند، توجه به نرخ True Positive و False Negative در تشخیص خروجی می‌باشد. علت اهمیت این موضوع را شرح دهید و بیان کنید که چه راهی برای ایجاد عدم توازن تاثیر خطا در این موارد پیشنهاد می‌کنید. حال روش پیشنهادی را برای رسیدن به خطای نزدیک به صفر در تشخیص سرطان پیاده کنید. بدست آمدن نتایج به دست آمده را با نتایج قسمت قبلی سوال مقایسه کنید.

توجه: استفاده از کتابخانه naïve bayes مجاز است.

گزارش: برنامه‌های نوشته شده به همراه نتایج خواسته شده در هر مرحله را ضمیمه کنید. همچنین یک بحث و بررسی در مورد اختلاف نتایج و مشخص کردن طبقه‌بندی‌کننده بهتر در این پروژه انجام دهید. گزارش شامل تحلیل و مقایسات و توضیحات کافی در مورد نحوه پیاده‌سازی‌ها، لیست برنامه‌ها و پارامترهای بکار رفته می‌باشد. لطفاً از نوشتن گزارش‌های طولانی اجتناب کنید.

توجه: کل تمرین و گزارش باید بصورت انفرادی نوشته شود، و به برنامه‌ها و گزارش‌های کپی شده نمره‌ای تعلق نمی‌گیرد.

موعد تحویل: ۲۲ دی ۱۴۰۱

روش تحویل: آپلود در courses

موفق باشید

سیدین