

یادگیری ماشین - نیمسال اول ۹۵-۱۳۹۴
تکلیف شماره ۲ - تحویل شنبه ۱۴/۹/۱۳۹۴

آشنایی با رگرسیون و دسته بندی خطی

۱- مجموعه داده tak2train.xlsx (که در سامانه موجود است) نمرات میان نیمسال، پایان نیمسال، تکلیف ۱، تکلیف ۲، تحقیق، پروژه، نمره نهائی، و وضعیت گذراندن یا نگذراندن درس تعدادی دانشجوی فرضی درس یادگیری ماشین را نشان می دهد. ستون اول ردیف هر دانشجو را نشان می دهد. توسط دستور xlsread() در نرم افزار متلب، این مقادیر را خوانده و در متغیر X قرار دهید. ستونی که مربوط به نمره نهائی است را در متغیر t قرار دهید و سپس ستونهای اول و آخر را از متغیر X حذف نمایید. در این حالت نمرات هر دانشجو در هر سطر X و نمره نهائی او در سطر متناظر آن در t قرار خواهد داشت. (دقت کنید که دستور xlsread() ستون آخر که مقادیر غیر عددی دارد را نمی خواند). مجموعه داده tak2test.xlsx (که در سامانه موجود است را نیز بصورت مشابه خوانده و برای مراحل آزمون آنرا نگهداری نمایید).

۲- در صورتی که بخواهیم از رگرسیون خطی برای تخمین نمره هر دانشجو بر اساس نمرات هر یک از مواد درسی استفاده نمائیم، با استفاده از روش بیشینه سازی درستمنائی و حداقل مربعات مقدار وزنهای این رگرسیون را بدست آورید بخش ۱-۳-۱). در این حالت درصدی که هر یک از مواد در تعیین نمره نهائی داشته است چه مقدار بوده است؟ با ۴ نمونه جدید توسط متلب پاسخهای خود را چک کنید.

۳- حال با روش نزول گرادیانی تصادفی مقادیر وزنها را بدست آورید. آیا مقادیر وزنهای بدست آمده با روشی که در بند ۲ بدست آمد یکسان است؟ نمودار خطا بر حسب تعداد دفعات تکرار را رسم نمایید. تأثیر نرخ یادگیری و تعداد دفعات تکرار در بدست آوردن مقدار وزنها چگونه است؟ با ۴ نمونه جدید توسط متلب پاسخهای خود را چک کنید.

۴- فرض کنید کسانی که نمره نهائی آنها بیش از یا برابر با ۱۲ است موفق به گذراندن درس شده (pass) و کسانی که کمتر از ۱۲ شده اند موفق به گذراندن درس نشده اند (fail). با روش حداقل مربعات دسته بندی خطی مجموعه داده گفته شده را انجام دهید. با ۴ نمونه جدید توسط متلب پاسخهای خود را چک کنید.

۵- حال با استفاده از روش رگرسیون لاجیستیک دسته بندی خطی را انجام دهید. با ۴ نمونه جدید توسط متلب پاسخهای خود را چک کنید.

۶- چگونگی برخورد با نقاط برون افتاده توسط روشهای فوق را بررسی کنید.

بارم بندی و نمره های این تکلیف فرضی می باشند و الزاماً مشابه بarm بندی واقعی نمی باشند. پاسخها به سؤالات فوق و نتایج آزمایشهای خود را در گزارشی به فرمت PDF به همراه برنامه های خود تحویل نمائید.

لطفاً به نکات زیر توجه نمائید:

- تکالیف تا دو روز بعد از موعد مقرر قابل تحویل می باشند و به ازاء هر روز تأخیر ۱۰٪ از نمره آن کسر می گردد.
- فقط به تکالیفی که در سامانه تحویل داده شوند نمره داده خواهد شد.
- به تکالیف مشابه نمره ای تعلق نخواهد گرفت.

موفق باشید

پالهنک