



به نام خدا

تمرینات سری دوم درس هوش محاسباتی

1400/03/13

یک کد پایتون ایجاد کنید و تمامی سوالات را به ترتیب در آن جواب دهید. تمامی نکات مورد نیاز برای سوال را به صورت کامنت بنویسید.

رگرسیون یک متغیره

- 1- ابتدا مجموعه داده با نام theCars را بخوانید. این مجموعه داده اطلاعاتی از 36 ماشین شامل نوع ماشین، مدل ماشین، حجم، وزن و مقدار گاز خروجی تولید شده (CO_2) توسط این ماشین‌ها، می‌باشد.
- 2- نمودار حجم- CO_2 را رسم کنید (محور x ، حجم و محور y میزان CO_2 را به صورت scatter نشان دهد).
- 3- نمودار وزن- CO_2 را رسم کنید.
- 4- نمودار سه بعدی حجم، وزن و CO_2 را رسم کنید (محور x ، حجم، محور y ، وزن و محور z میزان CO_2 را نشان دهد).
- 5- 36 نمونه را به دو دسته آموزش (30 نمونه) و تست (6 نمونه) تقسیم کنید. در این تقسیم بندی 6 نمونه Fiat-500، Hyundai-I20، Suzuki-Swift، Ford-Focus، Mazda-3 و Opel-Astra را به عنوان نمونه تست در نظر بگیرید. کد پایتون را به صورتی بنویسید که بر روی ستون‌های یک و دو جستجو کند و این 6 نمونه را پیدا کند و در دسته تست قرار دهد.
- 6- با استفاده از کتابخانه sk-learn خط برازش شده برای حجم- CO_2 را برای نمونه‌های آموزش بدست آورید. مقدار شیب و عرض از مبدا چقدر است؟
- 7- مقدار خطا برای نمونه‌های تست چقدر است؟ (مقدار خطا از میانگین مربع تفاضل بین مقدار تخمین زده شده و مقدار واقعی بدست می‌آید).
- 8- با استفاده از کدهای نوشته در کلاس برای رگرسیون خطی یک متغیره خط برازش شده برای حجم- CO_2 را بدست آورید. به ازای چه مقدار α (گام در الگوریتم نزول گرادیانی)، ضرایب به دست آمده به ضرایب سوال قبل نزدیکتر است؟
- 9- سوالات 6 تا 9 را برای وزن- CO_2 تکرار کنید.

رگرسیون چند متغیره

- 10- حال از دو متغیر حجم و وزن برای تخمین CO_2 برای داده‌های آموزش استفاده کنید. (از کتابخانه sk-learn استفاده کنید). مقدار خطا بر روی داده‌های تست چقدر است؟ توضیح دهید.
- 11- در این سوال هر 36 داده را در نظر بگیرید. با استفاده از مدل به دست آمده در سوال قبل، برای کدامیک از انواع ماشین (Car Categorical)، میانگین مقدار مربع خطا کمتر است؟

رگرسیون لاجستیک

12- حال سه متغیر حجم، وزن و CO2 را به عنوان ورودی در نظر بگیرید. هدف این است که با استفاده از این سه متغیر نوع ماشین (Car Categorical) را دسته بندی کنیم (بدون توجه به مدل ماشین ها). همانند قبل داده ها را به دو دسته آموزش و تست تقسیم کنید. مدل را بر روی داده های آموزش به دست آورید. (می توانید از کتابخانه sk-learn استفاده کنید).

13- مقدار دقت برای داده های آموزش چقدر است؟

14- مقدار دقت برای داده های تست چقدر است؟

15- در این سوال هر 36 داده را در نظر بگیرید. با استفاده از مدل به دست آمده در سوال 12 برای کدامیک از انواع ماشین ها (Car Categorical) مقدار دقت دسته بندی بیشتر از بقیه می باشد.

توجه:

- سوالات خود را در سامانه مودل مطرح کنید.
- تمامی فرضیاتی که در سوال ها در نظر می گیرید، به عنوان کامنت در برنامه های تان بنویسید.
- برنامه های خود در یک فایل ارسال کنید. سوالات را به ترتیب پاسخ دهید. در کامنت ها شماره سوال و در صورتی که نیاز بود به صورت انگلیسی توضیحات مربوط به سوال را بنویسید.
- مهلت تحویل تمرین سری اول تا تاریخ 28 خرداد ماه می باشد.
- لطفا تا قبل از برگزاری کلاس حل تمرین این سری از سوالات، جواب های خود را در اختیار دوستانتان قرار ندهید. به دانشجویان محترمی که جواب های مشابه ارسال کنند، نمره ای تعلق نخواهد گرفت.

موفق باشید. صادقی