

## تمارین یادگیری ماشین سری شماره یک

### معضل ابعاد:

#### پرسش ۱

فرض کنید دایره ای به قطر  $a$  درون یک مربع به طول  $a$  محاط شده است. چه مقدار از فضای مربع بلا استفاده مانده است؟ ابعاد دایره و مربع را یک بعد و سپس یک بعد دیگر افزایش دهید و برای همه حالات این حجم باقی مانده را محاسبه کنید. بررسی کنید این فضای اشغال شده با افزایش ابعاد به چه مقداری همگرا می شود ( با اثبات ) .

#### پرسش ۲

- فرض کنید که در فضای یک بعد،  $50$  نقطه تصادفی در بازه  $0$  تا  $100$  قرار دارند. این  $50$  نقطه را تولید کنید. سپس این نقاط را بر روی محور یک بعد که به  $5$  بازه مساوی تقسیم شده است، نمایش دهید.
- حال فرض کنید  $50$  نقطه تصادفی در فضای دو بعد قرار دارند که مقدار  $x$  و  $y$  آنها هر دو بین  $0$  تا  $100$  قرار دارد. اکنون، در صفحه مختصات در بازی  $0$  تا  $100$  هر محور، آن را به  $5$  بازه مساوی تقسیم کرده و نقاط را نمایش دهید.
- اکنون فرض کنید که  $50$  نقطه تصادفی در فضای سه بعدی داریم که مقدار متناظر هر بعد آنها عددی بین  $0$  تا  $100$  است. صفحه مختصات را در بازه ای که این نقاط قرار گرفته اند، به  $5$  بازه مساوی تقسیم کنید و سپس نقاط را نمایش دهید.
- با توجه به خروجی سه پرسش پیشین، چه تحلیلی به دست می آید؟ توضیح دهید. به علاوه برای هر سه پرسش قبل، چه کسری از نقاط در هر سلول ( از مختصاتی که تقسیم شده است؛ به طور مثال در سه بعد، این سلول به صورت مکعب است) قرار دارد؟ به صورت تئوری و عملی پاسخ دهید و نتیجه را تحلیل کنید.

#### پرسش ۳

- فرض کنید که یک دایره درون یک مربع قرار دارد. حال فرض کنید که هدف این است که داده های ما (دایره) به صورت استاندارد مقیاس شده اند، به این معنا که بر روی صفر مرکز یافته اند. در نتیجه مرکز دایره بر روی صفر قرار دارد. این دایره را رسم کنید. ( شعاع دایره یک است)
- برای پرسش قبل، کسر حجمی که دایره درون مربع اشغال کرده است را به دست آورید. به علاوه، همین پرسش را برای سه بعد و حالتی که داده به صورت کره ای درون یک مکعب است، بیابید و نمایش دهید.
- با توجه به نتایج پرسش های قبل، با تغییر بعد به چه نتیجه ای می توان دست یافت؟ تحلیل کنید.
- حال برای  $4$  بعد تا  $14$  بعد، نقاط رندومی را با توزیعی بین  $0$  و  $1$  در نظر بگیرید و فاصله ای این نقاط را از مرکز مختصات حساب کنید و این فاصله را به ازای هر بعد گزارش کنید. سپس میزان درصد حجمی که یک کره در فضای  $n$  بعدی درون یک مکعب در این بعد اشغال می کند حساب کرده و در یک نمودار نمایش دهید. نتیجه را تحلیل کنید.

#### پرسش ۴

- فرض کنید ۵۰ نقطه‌ی تصادفی با توزیع نرمال را در بازه‌ی ۰ تا ۱۰۰ در ابعاد ۱، ۳، ۱۰ و ۱۰۰ قرار دارند. میانگین فاصله‌ی ۵ نزدیک‌ترین همسایه‌ی ۱۰ نقطه‌ی تصادفی از این نقاط را در هر یک از ابعاد گفته شده به دست آورید و نتیجه را به طور کامل تحلیل کنید. چند معیار (metric) را نظیر L1, Cosine بررسی کنید و نتایج را مقایسه کنید. به صورت کلی آیا تغییر روش اندازه‌گیری می‌تواند برای معضل ابعاد روش درستی باشد؟ بدون بررسی عمیق جزئیات تحقیق کنید چه روش‌های دیگری برای رفع این مشکلات وجود دارد.

پرسش ۵

مثالهای مطرح شده در بخش 1.4 از کتاب مرجع [۱] را که در کلاس درسی به آنها اشاره شده است، به طور کامل توضیح دهید.

پرسش ۶

تمارین 1.15 و 1.16 از کتاب مرجع [۱] را حل کنید.

توجه: \* پرسش‌های ۲ و ۳ و ۴ را طبق فایل ipynb. \* ضمیمه شده، پیاده‌سازی کنید.

## مراجع

[1] Bishop, C. M. (2006). Pattern recognition and machine learning. New York :Springer,

دوره یادگیری ماشین / دانشگاه تربیت مدرس / گروه علوم کامپیوتر / نیمسال اول سال تحصیلی ۱۴۰۱-۱۴۰۲  
استاد درس : دکتر منصور رزقی‌آهق (Mrezghi.ir)