



سوال ۱: تحلیل و دسته‌بندی انواع شراب با KNN، PCA و Regression

هدف پروژه: در این پروژه شما با استفاده از دیتاست Wine به تحلیل و دسته‌بندی انواع شراب می‌پردازید. شما با تکنیک‌های مختلف یادگیری ماشین از جمله KNN، PCA، Rule-Based، Linear/Logistic Regression و اثر معیار فاصله و Split داده‌ها آشنا می‌شوید و عملکرد هر روش را مقایسه می‌کنید.

- **Wine Dataset** موجود در `sklearn.datasets`

- **ویژگی‌ها: (Features)** ۱۳ ویژگی عددی مربوط به ترکیبات شیمیایی شراب

- **کلاس‌ها: (Target)** ۳ نوع شراب مختلف (Class 0, 1, 2)

مرحله اول KNN ساده:

۱. داده‌ها را به دو بخش **Training** و **Testing** تقسیم کنید (مثلاً ۷۰٪ آموزش، ۳۰٪ تست).

۲. یک مدل **KNN** با مقدار $k=3$ بسازید.

۳. دقت مدل را روی داده‌های تست محاسبه کنید.

مرحله دوم: کاهش ابعاد با PCA

۱. داده‌ها را با **PCA** از ۱۳ ویژگی به ۲ مؤلفه اصلی کاهش دهید.

۲. مدل **KNN** را روی داده‌های کاهش‌یافته آموزش دهید و دقت آن را روی داده‌های تست محاسبه کنید.

مرحله سوم: سیستم Rule-Based

۱. یک سیستم ساده **Rule-Based** بر اساس یکی از ویژگی‌ها مثلاً Alcohol طراحی کنید:

$\text{Alcohol} < 12 \rightarrow \text{Class 0}$ ○

$12 < \text{Alcohol} < 13 \rightarrow \text{Class 1}$ ○

$\text{Alcohol} > 13 \rightarrow \text{Class 2}$ ○

۲. دقت **Rule-Based** را روی داده‌های **Test** محاسبه کنید و با **KNN** مقایسه کنید

مرحله چهارم: اثر معیارهای فاصله (Distance Metrics) و Decision Boundary

۱. مدل **KNN** با $k=3$ و معیارهای فاصله مختلف بسازید:

Mahalanobis - Chebyshev – Euclidean - Manhattan – Cosine

۲. دقت هر مدل را روی داده‌های تست محاسبه کنید.

۳. داده‌ها را با **PCA** به ۲ بعد کاهش دهید و **Decision Boundary** هر مدل را رسم کنید.

مرحله پنجم: Soft Encoding: Classification-as-Regression

۱. کلاس‌ها را به مقادیر پیوسته و ترتیبی تبدیل کنید:

○ $\text{Class } 0 \rightarrow 0$

○ $\text{Class } 1 \rightarrow 0.5$

○ $\text{Class } 2 \rightarrow 1$

۲. مدل **Linear Regression** آموزش دهید و خروجی پیوسته تولید کنید.

۳. خروجی مدل را به کلاس واقعی نزدیک‌ترین مقدار تبدیل کنید.

۴. دقت پیش‌بینی‌ها را روی داده‌های تست محاسبه کنید.

۵. داده‌ها را با PCA به ۲ بعد کاهش دهید و **Decision Boundary** را رسم کنید.

۶. تحلیل کنید:

○ آیا Soft Encoding باعث سوگیری مدل می‌شود؟

○ آیا مدل ممکن است کلاس میانی را بیش از حد بین کلاس‌های دیگر پیش‌بینی کند؟

توجه داشته باشید که Decision Boundary فقط برای داده‌های ۲ بعدی PCA رسم شود، ولی مدل‌ها روی داده‌های کامل ۱۳ بعدی آموزش داده شوند.

بخش دوم:

یکی از تاپیک‌های پیشنهادی را انتخاب کرده و در روز تحویل تمرین، ارائه دهید. هدف این ارائه‌ها، درک عمیق‌تر و تسلط بر مفاهیم ریاضیاتی و الگوریتم‌های مرتبط با هوش مصنوعی و یادگیری ماشین است.

تاپیک‌های پیشنهادی برای ارائه:

۱- ماتریس‌ها و عملیات ماتریسی:

- تعریف ماتریس: شناخت ماتریس‌ها، ابعاد و نوع آن‌ها (مربعی، مستطیلی)، عملیات ماتریسی (جمع، تفریق، ضرب ماتریس‌ها)

- ماتریس معکوس: مفهوم و کاربرد آن در حل سیستم‌های معادلات خطی.

- ماتریس‌های ویژه: مفهوم و کاربرد آن‌ها در PCA و دیگر الگوریتم‌ها.

- دترمینان، رنک و نرم ماتریس‌ها.

۲- بردارها و فضای برداری

- تعریف بردار: شناخت بردارها و نمایندگی آن‌ها در فضای n بعدی و عملیات بر روی بردارها (جمع، تفریق، ضرب اسکالر و ضرب داخلی).

- مقدار طول بردار و نرمال‌سازی: مفهوم و محاسبه طول بردار و نرمال‌سازی آن.

۳- مقادیر و بردارهای ویژه

-تعریف مقادیر ویژه و بردارهای ویژه: نحوه محاسبه و کاربرد آن‌ها در تجزیه و تحلیل داده‌ها.

- تجزیه ماتریس‌ها: مانند تجزیه SVD و کاربرد آن در PCA.

۴- آمار و احتمال

- توزیع‌های آماری: شناخت توزیع‌های نرمال و دیگر توزیع‌های مهم. P-value

- میانگین، واریانس و انحراف معیار: محاسبه و کاربرد آن‌ها در تحلیل داده‌ها.

۵- تحلیل تفکیک خطی (LDA):

- تعریف LDA و مقایسه آن با PCA.

- مراحل LDA: شامل محاسبه میانگین کلاس‌ها، محاسبه ماتریس‌های واریانس بین و درون کلاس و تحلیل تفکیک.

- کاربرد LDA در شناسایی الگو و طبقه‌بندی داده‌ها.

۶- تحلیل و مولفه‌های اصلی PCA

- ریاضیات PCA: مراحل مختلف PCA شامل مرکز کردن داده‌ها، محاسبه کوواریانس و تجزیه مقادیر ویژه.

- کاربرد PCA: در کاهش ابعاد و تصویرسازی داده‌ها.

۷- انواع نمودار در matplotlib و کاربرد هر کدام.

بررسی نمودارهای مختلف مانند نمودار خطی، ستونی، پراکندگی، هیستوگرام و... و کاربرد آن‌ها در تحلیل داده‌ها

توجه:

- لطفاً توجه داشته باشید که هر یک از شما باید یکی از این تایپیک‌ها را انتخاب کند و از انتخاب موضوعات تکراری در هر گروه خودداری کنید. بنابراین، قبل از انتخاب، با یکدیگر در گروه خود هماهنگ کنید تا از تکرار موضوعات جلوگیری شود.

-زمان ارائه برای هر نفر بین ۸ الی ۱۰ دقیقه خواهد بود.

- سعی کنید در ارائه خود، مفاهیم را به طور واضح توضیح دهید و مثال‌های عملی را در صورت امکان شامل کنید.