

Load Wine

این کد یک برنامه یادگیری ماشین است که از الگوریتم **K-Nearest Neighbors (KNN)** برای طبقه‌بندی داده‌های مجموعه‌داده شراب (**Wine Dataset**) استفاده می‌کند. هدف اصلی این برنامه بررسی تأثیر **متریک‌های مختلف فاصله** بر دقت مدل KNN است. در ادامه، هر بخش از کد به طور کامل و با جزئیات توضیح داده می‌شود:

۱. وارد کردن کتابخانه‌های لازم

- `numpy`: برای انجام محاسبات عددی و کار با آرایه‌ها.
- `matplotlib.pyplot`: برای رسم نمودارها.
- `sklearn.datasets.load_wine`: برای بارگذاری مجموعه‌داده شراب.
- `sklearn.model_selection.train_test_split`: برای تقسیم داده‌ها به مجموعه‌های آموزشی و تست.
- `sklearn.metrics.accuracy_score`: برای محاسبه دقت مدل.
- `sklearn.neighbors.KNeighborsClassifier`: برای پیاده‌سازی مدل KNN.

۲. بارگذاری داده‌های شراب

- مجموعه‌داده شراب (`load_wine`) بارگذاری می‌شود. این مجموعه‌داده شامل ۱۳ ویژگی (مانند اسیدیته، الکل، فلاونوئیدها و ...) و ۳ کلاس (نوع شراب) است.

۳. تقسیم داده‌ها برای نسبت‌های مختلف

- داده‌ها به نسبت‌های مختلف آموزشی و تست تقسیم می‌شوند. این کار در یک حلقه انجام می‌شود:
 - `num = 20`: تعداد تکرارها برای تقسیم داده‌ها.
 - `test_size`: اندازه مجموعه تست به صورت درصدی از کل داده‌ها (از ۹٪ تا ۲۸٪ با گام ۱٪).
 - `x_train_list, x_test_list, y_train_list, y_test_list`: لیست‌هایی برای ذخیره داده‌های آموزشی و تست در هر تکرار.
 - `size_train, size_test`: لیست‌هایی برای ذخیره اندازه‌های مجموعه‌های آموزشی و تست.

۴. لیست متریک‌های مورد نظر

- لیستی از متریک‌های فاصله مختلف تعریف می‌شود:
 - `euclidean`: فاصله اقلیدسی.
 - `manhattan`: فاصله منهتن.
 - `chebyshev`: فاصله چبیشف.
 - `minkowski`: فاصله مینکوفسکی (با $p=3$).
 - `cosine`: فاصله کسینوسی.
 - `correlation`: فاصله همبستگی.
 - `hamming`: فاصله هامینگ.

۵. ترسیم دقت برای هر متریک

- برای هر متریک، دقت مدل KNN محاسبه و نمودار آن رسم می‌شود. مراحل آن به شرح زیر است:
 ۱. حلقه روی متریک‌ها:
 - برای هر متریک در لیست `metrics`، دقت مدل KNN محاسبه می‌شود.
 ۲. حلقه روی تقسیم‌های داده:
 - برای هر تقسیم داده (از ۱ تا ۲۰)، مدل KNN با متریک فعلی آموزش داده می‌شود.
 - پیش‌بینی‌ها انجام می‌شوند و دقت مدل محاسبه می‌شود.
 - دقت‌ها در لیست `accuracies` ذخیره می‌شوند.
 ۳. رسم نمودار:
 - برای هر متریک، یک زیرنمودار (subplot) رسم می‌شود.
 - نمودار دقت بر حسب تعداد تست‌ها نمایش داده می‌شود.
 - عنوان نمودار شامل نام متریک است.
 - محور `x` نشان‌دهنده تعداد تست‌ها و محور `y` نشان‌دهنده دقت است.
 - از `plt.grid()` برای نمایش خطوط شبکه‌ای استفاده می‌شود.
 - محدوده محور `y` بین ۰ تا ۱ تنظیم می‌شود.

۶. تنظیمات نمودار

- `plt.figure(figsize=(15, 12))`: اندازه کلی نمودار تنظیم می‌شود.
- `plt.subplot(4, 2, idx + 1)`: برای هر متریک، یک زیرنمودار در یک شبکه 2×4 ایجاد می‌شود.
- `plt.tight_layout()`: برای جلوگیری از همپوشانی زیرنمودارها استفاده می‌شود.
- `plt.show()`: نمودار نهایی نمایش داده می‌شود.

۷. اجرای KNN با متریک‌های مختلف

- برای هر متریک، مدل KNN با پارامترهای زیر اجرا می‌شود:
 - `n_neighbors=3`: تعداد همسایه‌ها برابر با ۳ است.
 - `metric=metric`: متریک فاصله مورد استفاده در مدل.

۸. محاسبه دقت

- دقت مدل با استفاده از `accuracy_score` محاسبه می‌شود. این تابع پیش‌بینی‌های مدل (`y_pred`) را با برچسب‌های واقعی (`y_test`) مقایسه می‌کند.

۹. نمایش نتایج

- برای هر متریک، یک نمودار جداگانه رسم می‌شود که دقت مدل را در هر تقسیم داده نشان می‌دهد.
- نمودارها به صورت زیرنمودارهای مجزا در یک پنجره نمایش داده می‌شوند.

۱۰. خروجی برنامه

- **نمودارها:** برای هر متریک، یک نمودار نمایش داده می‌شود که دقت مدل KNN را در هر تقسیم داده نشان می‌دهد.
- **مقایسه متریک‌ها:** با مقایسه نمودارها، می‌توان فهمید که کدام متریک بهترین عملکرد را دارد.

مثال خروجی

نمودارهای زیر برای هر متریک نمایش داده می‌شوند:

- ۱. **Accuracy By Metric: euclidean**
- ۲. **Accuracy By Metric: manhattan**
- ۳. **Accuracy By Metric: chebyshev**
- ۴. **Accuracy By Metric: minkowski**
- ۵. **Accuracy By Metric: cosine**
- ۶. **Accuracy By Metric: correlation**
- ۷. **Accuracy By Metric: hamming**

هر نمودار دقت مدل KNN را برای ۲۰ تقسیم داده مختلف نشان می‌دهد.

جمع‌بندی

این برنامه به طور سیستماتیک تأثیر متریک‌های مختلف فاصله بر دقت مدل KNN را بررسی می‌کند. با استفاده از این برنامه، می‌توان فهمید که:

- کدام متریک بهترین عملکرد را دارد.
- چگونه دقت مدل با تغییر اندازه مجموعه تست تغییر می‌کند.
- آیا برخی متریک‌ها برای این مجموعه داده مناسب‌تر هستند یا خیر.