

Load Wine

تعریف تابع `calculate_accuracy`

این تابع دقت مدل KNN را محاسبه می‌کند. مراحل آن به طور کامل توضیح داده می‌شود:

• ورودی‌ها:

- `dataset`: مجموعه داده ورودی به صورت دیتافریم.
- `test_size`: اندازه مجموعه تست به صورت درصدی از کل داده‌ها (مثلاً ۰.۲ به معنای ۲۰٪).
- `features_to_use`: لیست اندیس‌های ویژگی‌هایی که باید در مدل استفاده شوند.

• مراحل اجرا:

۱. جداسازی ویژگی‌ها و برچسب‌ها:
 - `X`: ویژگی‌های مجموعه داده (بدون ستون `target`).
 - `y`: برچسب‌های مجموعه داده (ستون `target`).
۲. تقسیم داده‌ها به مجموعه‌های آموزشی و تست:
 - داده‌ها به دو بخش `train` و `test` تقسیم می‌شوند. اندازه مجموعه تست بر اساس ورودی `test_size` تعیین می‌شود.
۳. استانداردسازی ویژگی‌ها:
 - ویژگی‌ها با استفاده از `StandardScaler` استانداردسازی می‌شوند. این کار برای اطمینان از این است که همه ویژگی‌ها در یک مقیاس قرار دارند و هیچ ویژگی به دلیل مقادیر بزرگ‌تر، تأثیر بیشتری بر مدل نداشته باشد.
۴. محاسبه تعداد همسایه‌ها (`k`):
 - تعداد همسایه‌ها (`k`) به عنوان جذر تعداد نمونه‌های آموزشی محاسبه می‌شود. این یک روش رایج برای انتخاب `k` در الگوریتم KNN است.
۵. آموزش مدل KNN و پیش‌بینی:
 - برای هر ویژگی مشخص شده (`features_to_use`)، مدل KNN آموزش داده می‌شود و پیش‌بینی‌ها انجام می‌شوند.
 - پیش‌بینی‌ها برای هر ویژگی در لیست `predicted_labels_per_features` ذخیره می‌شوند.
۶. ترکیب پیش‌بینی‌ها:
 - پیش‌بینی‌های تمام ویژگی‌ها با هم ترکیب شده و مُد (مقدار پرتکرار) به عنوان پیش‌بینی نهایی انتخاب می‌شود. این کار با استفاده از `scipy.stats.mode` انجام می‌شود.
۷. محاسبه دقت:
 - دقت مدل (`overall_accuracy`) با مقایسه پیش‌بینی‌های نهایی با برچسب‌های واقعی (`y_test`) محاسبه می‌شود.
 - همچنین، دقت مدل زمانی که از تمام ویژگی‌ها استفاده می‌شود (`accuracy1`) نیز محاسبه می‌شود.

• خروجی:

- `accuracy1`: دقت مدل زمانی که از تمام ویژگی‌ها استفاده می‌شود.
- `overall_accuracy`: دقت مدل زمانی که از ویژگی‌های انتخابی استفاده می‌شود.

۳. بارگذاری مجموعه داده شراب

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

- داده‌ها به یک دیتافریم (df) تبدیل می‌شوند تا پردازش آن‌ها آسان‌تر شود.

۴. تعریف تابع run_knn

این تابع مدل KNN را برای تعداد مشخصی از تکرارها اجرا می‌کند. مراحل آن به شرح زیر است:

- **ورودی‌ها:**
 - iterations: تعداد تکرارهایی که مدل باید اجرا شود.
- **مراحل اجرا:**
 ۱. **تکرار مدل:**
 - در هر تکرار، اندازه مجموعه تست (test_size) به صورت تصادفی بین ۱۰٪ تا ۳۰٪ انتخاب می‌شود.
 - تابع calculate_accuracy فراخوانی می‌شود و دقت مدل محاسبه می‌شود.
 ۲. **شمارش ویژگی‌های مهم:**
 - اگر دقت مدل در یک تکرار بیشتر از ۷۰٪ باشد، ویژگی‌هایی که در آن تکرار استفاده شده‌اند، شمارش می‌شوند.
- **خروجی:**
 - all_accuracy_results: لیستی از دقت‌های مدل در تمام تکرارها.
 - feature_counts: تعداد دفعاتی که هر ویژگی در تکرارهایی با دقت بالای ۷۰٪ استفاده شده است.

۵. دریافت تعداد تکرار از کاربر

- کاربر تعداد تکرارها را وارد می‌کند. این تعداد تعیین می‌کند که مدل چند بار اجرا شود.

۶. محاسبه دقت برای اندازه‌های مختلف مجموعه تست

- برنامه دقت مدل را برای اندازه‌های مختلف مجموعه تست (از ۱۱٪ تا ۲۵٪ با گام ۲٪) محاسبه می‌کند.
- برای هر اندازه تست، دقت مدل زمانی که از تمام ویژگی‌ها استفاده می‌شود (accuracy1) و زمانی که از ویژگی‌های انتخابی استفاده می‌شود (overall_accuracy) محاسبه می‌شود.

۷. رسم نمودار

- نتایج دقت برای اندازه‌های مختلف مجموعه تست در یک نمودار رسم می‌شوند.
- دو خط در نمودار وجود دارد:
 ۱. **خط آبی:** دقت مدل زمانی که از تمام ویژگی‌ها استفاده می‌شود. (accuracy1)
 ۲. **خط نارنجی:** دقت مدل زمانی که از ویژگی‌های انتخابی استفاده می‌شود. (overall_accuracy)
- نمودار به کاربر کمک می‌کند تا بفهمد چگونه اندازه مجموعه تست بر دقت مدل تأثیر می‌گذارد.

۸. خروجی برنامه

- **نمودار:** نموداری که دقت مدل را برای اندازه‌های مختلف مجموعه تست نشان می‌دهد.
 - **ویژگی‌های مهم:** ویژگی‌هایی که بیشترین تأثیر را بر دقت مدل دارند (ویژگی‌هایی که در تکرارهایی با دقت بالای ۷۰٪ استفاده شده‌اند).
-

جمع‌بندی

این برنامه به طور سیستماتیک دقت مدل KNN را برای مجموعه داده شراب بررسی می‌کند. با استفاده از این برنامه، می‌توان فهمید که:

- کدام ویژگی‌ها بیشترین تأثیر را بر دقت مدل دارند.
- چگونه اندازه مجموعه تست بر دقت مدل تأثیر می‌گذارد.
- آیا استفاده از زیرمجموعه‌ای از ویژگی‌ها می‌تواند دقت مدل را بهبود بخشد یا خیر.