

Wprowadzenie do uczenia maszynowego

Lista 5

Janusz Szwabiński

1. **Wprowadzenie do SVM:** Zaimplementuj SVM od podstaw dla danych, które są liniowo separowalne. Skup się na znalezieniu optymalnej hiperpłaszczyzny, która maksymalizuje margines między klasami.
2. **Zastosowanie SVM z scikit-learn:** Wczytaj zbiór danych **Breast Cancer** z biblioteki **scikit-learn**. Podziel dane na zbiór treningowy i testowy. Zbuduj model SVM z jądrem liniowym (`kernel='linear'`) i oceń jego wydajność.
3. **Jądro (kernel trick):** Wygeneruj syntetyczny zbiór danych, który nie jest liniowo separowalny (np. dane w kształcie okręgów). Wytrenuj model SVM z jądrem Gaussa (`kernel='rbf'`) i zwizualizuj granice decyzyjne. Porównaj wyniki z modelem SVM z jądrem liniowym i wyjaśnij, dlaczego jądro RBF jest w tym przypadku skuteczniejsze.
4. **Dostrajanie hiperparametrów SVM:** Dla modelu SVM na danych **Breast Cancer**, użyj **GridSearchCV** lub **RandomizedSearchCV** do optymalizacji hiperparametrów, takich jak parametr regularyzacji C i gamma (γ) dla jądra RBF. Porównaj wyniki z modelem o domyślnych parametrach.
5. **Porównanie XGBoost i LightGBM:** Zainstaluj i użyj bibliotek **xgboost** i **lightgbm**. Wytrenuj oba modele na danych **Breast Cancer** i porównaj ich czas treningu oraz dokładność. Zidentyfikuj, który z nich jest szybszy i dlaczego.
6. **Walidacja krzyżowa (Cross-validation):** Dla wybranego modelu (np. **XGBoost**) przeprowadź walidację krzyżową (**KFold** lub **StratifiedKFold**) w celu uzyskania bardziej rzetelnej oceny jego wydajności. Omów, dlaczego walidacja krzyżowa jest lepsza niż jednorazowy podział na zbiór treningowy i testowy.