

Wprowadzenie do uczenia maszynowego

Lista 6

Janusz Szwabiński

1. **Wprowadzenie do PCA:** Wygeneruj syntetyczny zbiór danych w 3D, w którym punkty układają się w linii lub płaszczyźnie. Użyj klasy PCA z `scikit-learn`, aby zredukować wymiarowość do 2D. Zwizualizuj dane przed i po redukcji.
2. **Analiza wariancji:** Dla zbioru danych `Iris` lub `Wine` użyj PCA i wygeneruj wykres wariancji wyjaśnionej. Zinterpretuj, ile głównych składowych należy zachować, aby wyjaśnić 95% całkowitej wariancji.
3. **Wizualizacja wielowymiarowych danych:** Zastosuj PCA do zbioru danych `Iris` lub `Wine`, redukując go do 2 głównych składowych. Stwórz wykres punktowy w 2D, kolorując punkty według ich klasy. Wyjaśnij, czy PCA jest skuteczną metodą do wizualizacji w tym przypadku.
4. **PCA jako pre-processing:** Wytrenuj model klasyfikacyjny (np. Regresję Logistyczną lub SVM) na danych `Wine` bez redukcji wymiarowości. Następnie, zastosuj PCA, redukując dane do 2 głównych składowych, i ponownie wytrenuj ten sam model. Porównaj dokładność i czas treningu w obu przypadkach.
5. **Liniowa Analiza Dyskryminacyjna (LDA):** Wytrenuj model LDA (`LinearDiscriminantAnalysis` z `scikit-learn`) na zbiorze danych `Wine`. Zredukuj dane do 2 wymiarów i zwizualizuj je. Porównaj wyniki z wizualizacją uzyskaną za pomocą PCA i omów różnice w podejściu.
6. **Zastosowanie na danych rzeczywistych:** Wczytaj zbiór danych do rozpoznawania twarzy (LFW - `Labeled Faces in the Wild`, dostępny w `scikit-learn`). Zastosuj PCA, aby zredukować wymiarowość obrazów. Wyświetl oryginalne obrazy i ich zrekonstruowane wersje po redukcji i dekompresji.