

Eksploracja danych

Piotr Lipiński

Lista zadań nr 8 – PCA, KPCA, LLE i inne

Zadanie 1. (4 punkty)

W pakiecie SciKit do Pythona dostępnych jest kilka popularnych metod redukcji wymiarowości i *manifold learning* (czyli metod wyznaczania rozmaitości różniczkowalnej, na której znajdują się rozpatrywane dane). Zapoznaj się z nimi wykonując skrypt umieszczony w materiałach do wykładu. Sprawdź jaki wpływ na działanie tych metod mają ich parametry.

Zadanie 2. (4 punkty)

Porównaj skuteczność klasyfikatora KNN działającego na oryginalnym zestawie danych zawierającym obrazy cyfr pisanych ręcznie (użytym w skrypcie z poprzedniego zadania) ze skutecznością klasyfikatorów KNN działających na zestawach danych zredukowanych do dwóch wymiarów za pomocą metod PCA, KPCA, Isomap, LLE, MDS, t-SNE.

Zadanie 3. (4 punkty)

Zastosuj metody redukcji wymiarowości i *manifold learning* z pierwszego zadania i spróbuj narysować na płaszczyźnie twarze ze zbioru danych AR Face Database rozpatrywanego w zadaniach z listy 7.