

Skalowanie i Wizualizacja Danych - MDS

Paweł Pudelski 127309, ISWD

Maj 2020

1 Wstęp

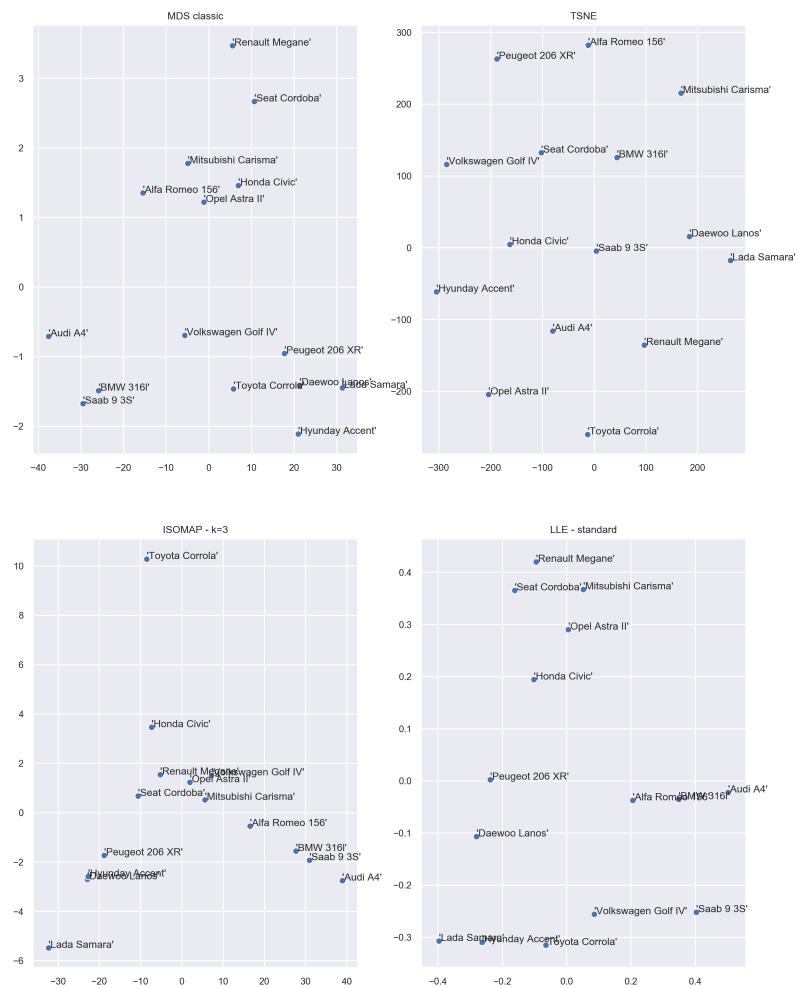
W sprawozdaniu przedstawiono wyniki redukcji wymiarowości w przypadku pozostawienia jedynie dwóch wymiarów. Wykorzystano tutaj następujące algorytmy:

- MDS w wersji klasycznej
- ISOMAP dla wielkości sąsiedztwa 3
- t-SNE
- LLE w wersji standardowej

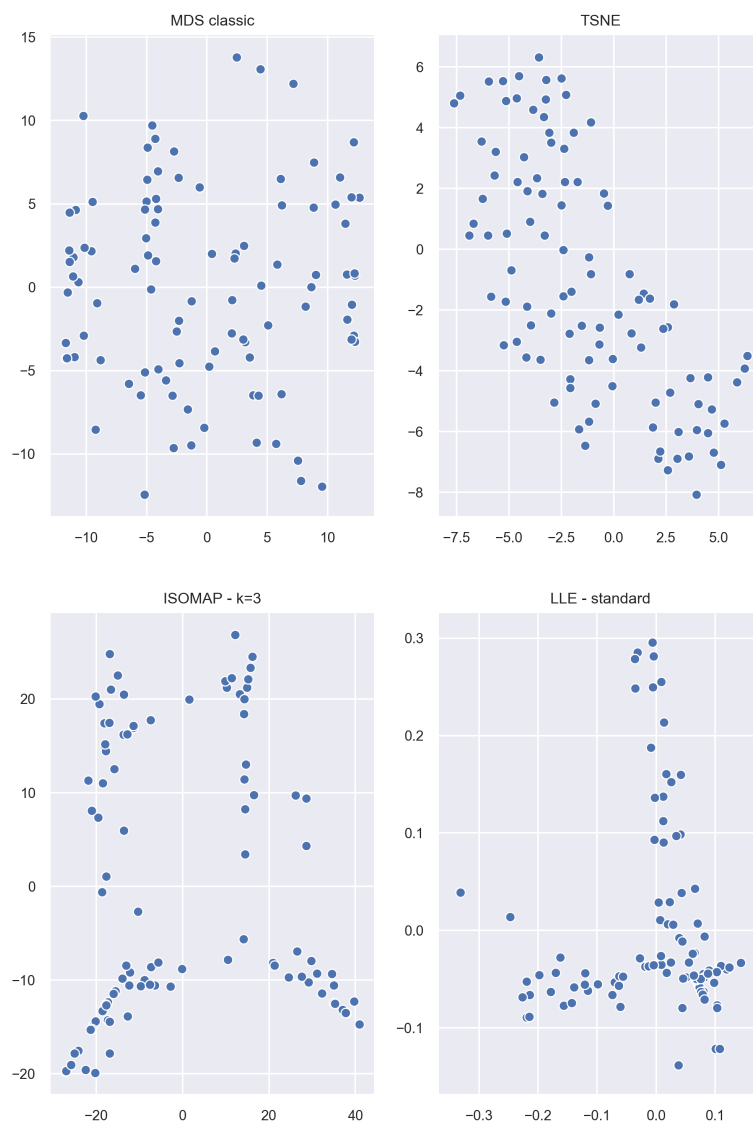
Trzeci dodatkowo wybrany zbiór (*Mordhau Weapons*) zawiera statystyki różnych broni białych pochodzących z gry Mordhau (wcielamy się tam w średnio-wiecznego rycerza i staczamy pojedynki z innymi graczami). Dla każdej z broni zostały opisane takie rzeczy jak obrażenia zadawane w głowę, tułów oraz nogi przeciwnikom bez zbroi, oraz z lekką, średnią i ciężką zbroją. Dodatkowo również przedstawiono tam koszt wytrzymałość w przypadku chybienia, finty (2 rodzaje) oraz samego zamach. Ostatnim atrybutem jest długość każdej z broni. Razem jest to 17 atrybutów liczbowych.

2 Wyniki

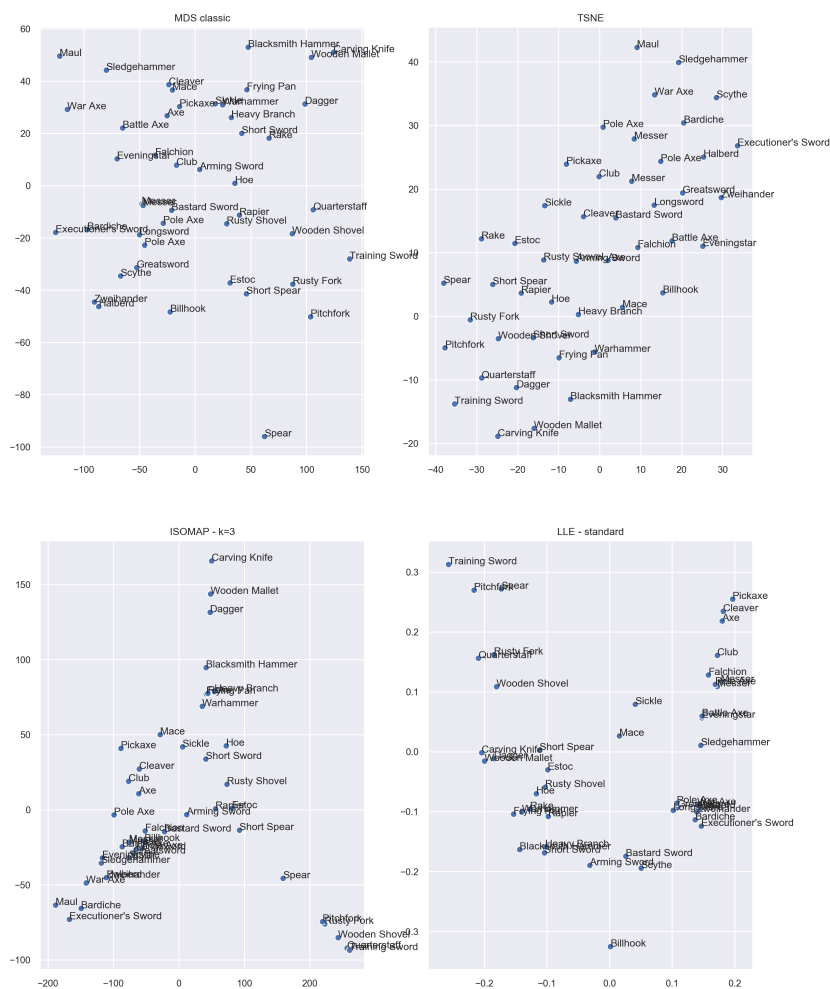
Rysunek 1: Wykres przedstawiający wyniki redukcji wymiarowości dla zbioru *Cars*.



Rysunek 2: Wykres przedstawiający wyniki redukcji wymiarowości dla zbioru *SwissRoll*.



Rysunek 3: Wykres przedstawiający wyniki redukcji wymiarowości dla zbioru *Mordhau Weapons*.



W przypadku algorytmu t-SNE widać, że stara się bardzo równomiernie rozłożyć wszystkie obserwacje w przestrzeni, tak aby sąsiadujące punkty miały równą odległości od siebie. Wyniki są bardzo podobnie niezależnie od liczby atrybutów, którymi opisanie były przykłady. Zdecydowanie reprezentacja jest tutaj

bardzo czytelna jednak słabo odzwierciedla to jak bardzo różnią się konkretne przykłady.

Algorytm ISO przy liczbie sąsiedztwa równej 3 widać, że próbuje dzielić przykłady właśnie na trzy oddzielne grupy. Przejawia się to po dość charakterystycznym ułożeniu punktów w kształt rozgwiazdy z trzema ramionami. Najwidoczniej parametr określający rozmiar sąsiedztwa w tym przypadku jest źle dobrany, gdyż po sprawdzeniu większych wartości punkty układają się w sposób bardziej zbliżony do MDS. Świadczy to dużej istotności odpowiedniego dobrania sąsiedztwa.

LLE natomiast stara się bardziej grupować przykłady i jednocześnie oddzielać powstałe grupy od siebie. Można powiedzieć, że działa podobnie do algorytmów analizy skupień. Jest to bardziej zauważalne dla danych z większą liczbą atrybutów, gdyż na ostatnim zbiorze widać, że powstają pewnie skupiska złożone z wielu punktów.

Klasyczne MDA opiera się na odległościach, z tego względu rozmieszczenie przykładów będzie bardzo zależeć od ich wyliczenia, na co wpływ ma również sam dobór metryki. Metoda to wydaje się bardzo dobrze prezentować stopień odmienności poszczególnych przykładów w przeciwieństwie do t-SNE.

Pod względem podobieństwa rozmieszczeń przykładów MDS jest dość blisko z LLE, co widać po *Renault Megane* i *Seat Cordoba* oraz parze *Audi A4* i *BMW 361i*. Podobnie jest również porównując MDS z ISOMAP patrząc na grupę: *Audi A4*, *BMW 361i*, *Saab 9 3S*, jednak tutaj *Toyota Corrola* została wyrzucona całkowicie na bok