

Nieliniowa redukcja wymiarowości

Marcin Mrugas 122580

Czerwiec 2019

1 Wstęp

Moim zadaniem było przeprowadzenie szeregu eksperymentów związanych z wizualizacją danych.

2 Metody

Wykorzystałem następujące metody:

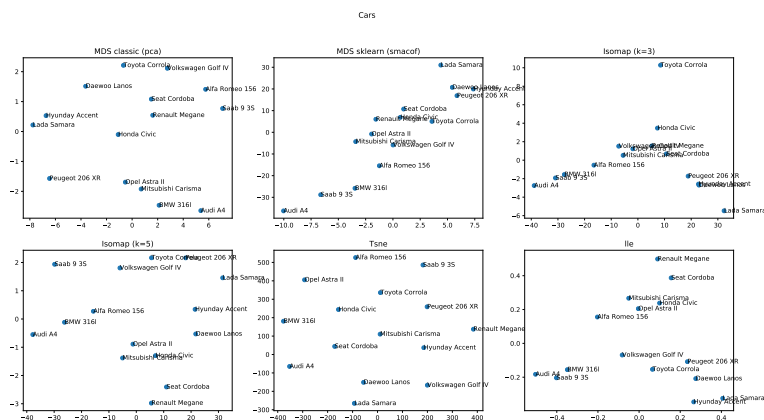
- Własna implementacja metody MDS oparta na PCA
- MDS z pakietu sklearn
- Isomap z $k=3$
- Isomap z $k=5$
- TSNE
- LocallyLinearEmbedding z pakietu sklearn

3 Samochody

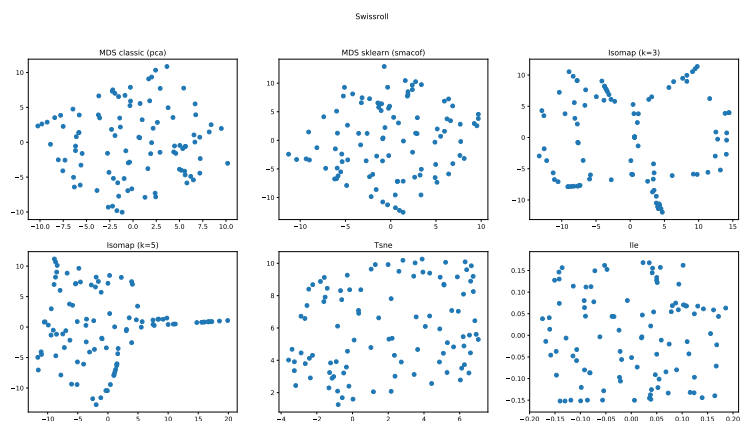
Porównując wyniki, najgorzej poradził sobie algorytm MDS z pakietu sklearn, w którym obiekty są mocno skorelowane.

4 Swissroll

We wszystkich algorytmach wizualizacja osiągnęła podobne efekty. Jednak pewne problemy są widoczne na wizualizacjach Isomap.



Rysunek 1: Wizualizacja zbioru danych cars



Rysunek 2: Wizualizacja zbioru danych swissroll

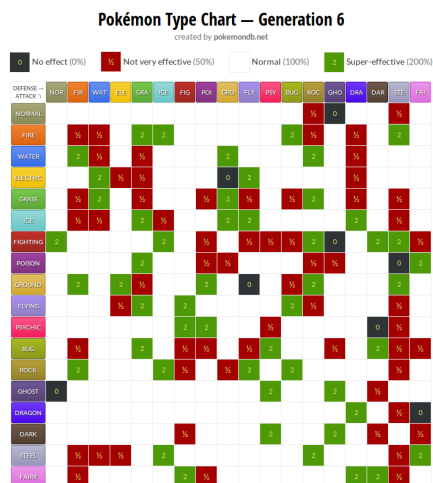
5 Własny zbiór danych

Jako własny zbiór danych wykorzystałem informacje pochodzące z konkursu Kaggle dotyczącym Pokemonów. Wybrałem tylko pokemony z pierwszej generacji, aby móc się odnieść do wyników ze względu na ich znajomość oraz mając na celu czytelność wykresów.

Ponieważ liczenie odległości pomiędzy typami, które były podane jako łańcuchy znaków, było kłopotliwe i nie ma jednokowego przydziału typów na skali porządkowej, zamieniłem dwie kolumny typów na odpowiadające im wiersze z

Tabela 1: Próbką danych o Pokemonach

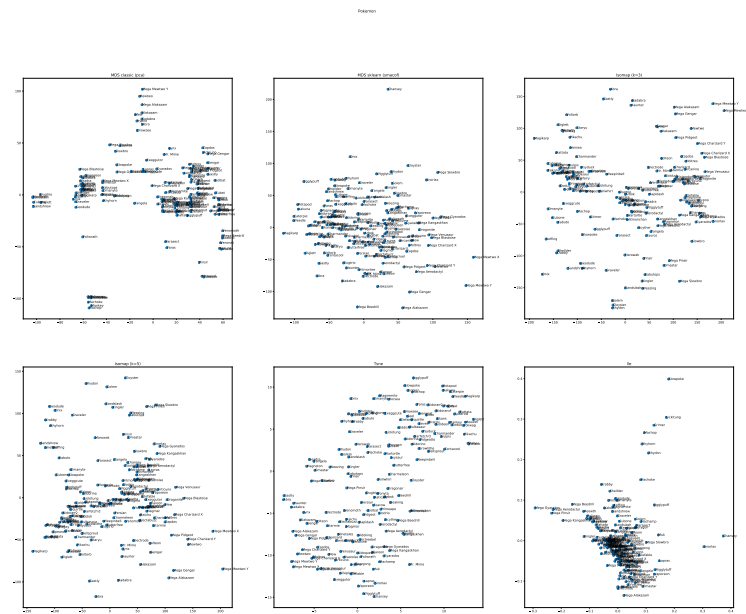
#	Name	Type 1	Type 2	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary
1	Bulbasaur	Grass	Poison	45	49	49	65	65	45	1	False
2	Ivysaur	Grass	Poison	60	62	63	80	80	60	1	False
3	Venusaur	Grass	Poison	80	82	83	100	100	80	1	False



Rysunek 3: Tabela przedstawiająca zależności między typami

tabeli 3 sił i słabości wobec innych typów. Motywacją było podobieństwo dwóch typów poprzez trudności z pokonaniem grup przeciwników tego samego typu. Takim sposobem typ *water* i *ground* powinny być blisko, ponieważ pokemony tego typu dobrze sobie radzą z przeciwnikami typu *fire* i *rock*, ale mają problemy z typem *grass*. Dodatkowo zmieniłem atrybut boolowski *Legendary* na wartości 0 i 1. Tak zmodyfikowane dane pozwoliły na wizualizację podobieństwa między pokemonami.

Dla wszystkich algorytmów dane zostały poprawnie odzorowane. Można zauważyć, że kolejne ewolucje pokemonów są blisko siebie i podobne typy pokemonów zostały rozmieszczone w swojej okolicy. I tak na przykład Gengar znajduje się blisko Haunter'a i Gastly'iego. Zaś wszystkie pokemony z przedrostkami Mega są bliżej jednej ze stron zwizualizowanych danych, co wcale nie było wprost zaznaczone w danych. Najlepiej moim zdaniem sprawdził się algorytm TSNE, który dodatkowo rozprzeczował obiekty równomiernie.



Rysunek 4: Wizualizacja podobieństw między pokemonami

Literatura

- [1] T7. Pokemon- Weedle's Cave. <https://www.kaggle.com/terminus7/pokemon-challengecombats.csv>.
- [2] zonation. pokemon-chart. <https://github.com/zonation/pokemon-chart>.