

# Podstawy uczenia maszynowego

## Główne składowe – raport

Mateusz Kocot

23 kwietnia 2021

### Spis treści

<b>1</b>	<b>Zbiór danych</b>	<b>1</b>
<b>2</b>	<b>PCA</b>	<b>2</b>
2.1	Średnie zdjęcie . . . . .	2
2.2	Nowe wektory bazowe . . . . .	2
2.3	Wariancja . . . . .	2
2.4	Redukcja wymiarowości . . . . .	3
2.5	Redukcja do dwóch wymiarów . . . . .	5

## 1 Zbiór danych

Zbiór danych przygotowano samodzielnie. Wybrano 6 ubrań reprezentujących klasy:

- marynarka,
- jeansy,
- szorty,
- bluza,
- koszulka,
- kurtka zimowa.



**Rys. 1:** Próbką zbioru danych

Wykonano 6 zdjęć każdego ubrania, co w sumie dało 36 zdjęć w zbiorze.

Zdjęcia przekonwertowano do rozmiaru  $128 \times 128$  oraz zapisano w skali szarości. Na rys. 1 przedstawiono próbkę zbioru danych – po jednym zdjęciu z każdej klasy.

## 2 PCA

### 2.1 Średnie zdjęcie

Przed wykonaniem transformacji PCA, zbiór został wycentrowany. Na rys. 2 przedstawiono średnie zdjęcie, które zostało odjęte od pozostałych. Widać na nim elementy (poświaty) wszystkich klas ze zbioru danych.



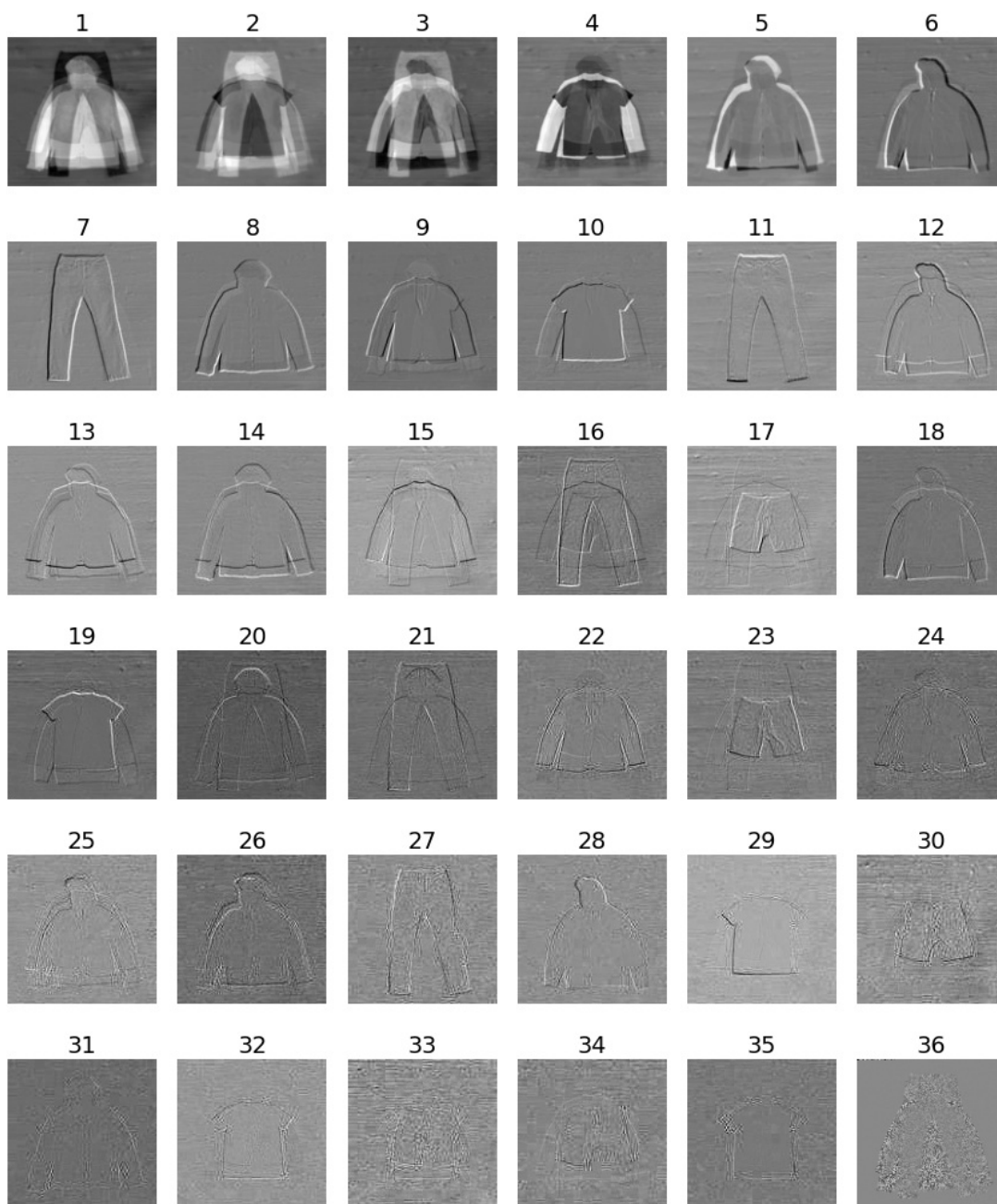
**Rys. 2:** Średnie zdjęcie

### 2.2 Nowe wektory bazowe

Następnie wykonano transformację PCA. Nowe wektory bazowe przedstawiono na rys. 3. Widać, że pierwsze skupiają się na najważniejszych cechach (niekoniecznie pojedynczych klas). Dalsze wektory odpowiadają za coraz bardziej szczegółowe cechy.

### 2.3 Wariancja

Wariancję (explained variance ratio) wektorów bazowych przedstawiono na rys 4. Jak widać, wektory bazowe od szóstego nie wnoszą już wiele informacji.

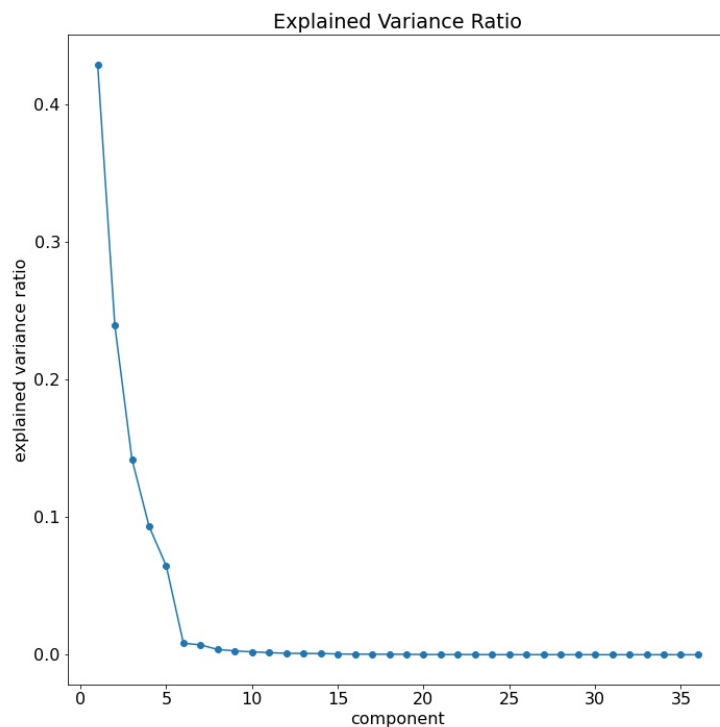


**Rys. 3:** Nowe wektory bazowe

## 2.4 Redukcja wymiarowości

Zredukowano wymiarowość do 4 najważniejszych cech, a następnie zredukowane obserwacje przekształcono z powrotem na fotografię (uwzględniając dodanie średniego zdjęcia) (rys. 5). Podobny proces przeprowadzono dla 16 wymiarów (rys. 6).

Przy 4 wymiarach, zdjęcia są rozpoznawalne, lecz widać, że pewne cechy zostały usunięte. Zdjęcia po 16 wymiarowym PCA są już niemal takie same jak ich pierwowzory.



**Rys. 4:** Wariancja wektorów bazowych (explained variance ratio)



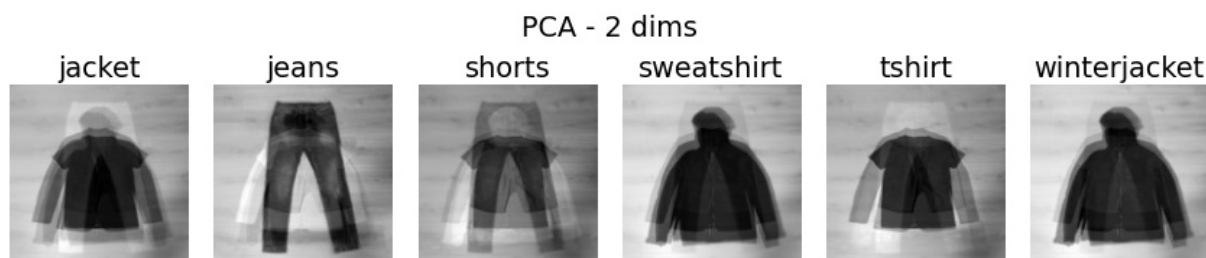
**Rys. 5:** Próbkę obserwacji po zredukowaniu do 4 wymiarów i przekształceniu z powrotem na fotografię.



**Rys. 6:** Próbkę obserwacji po zredukowaniu do 16 wymiarów i przekształceniu z powrotem na fotografię.

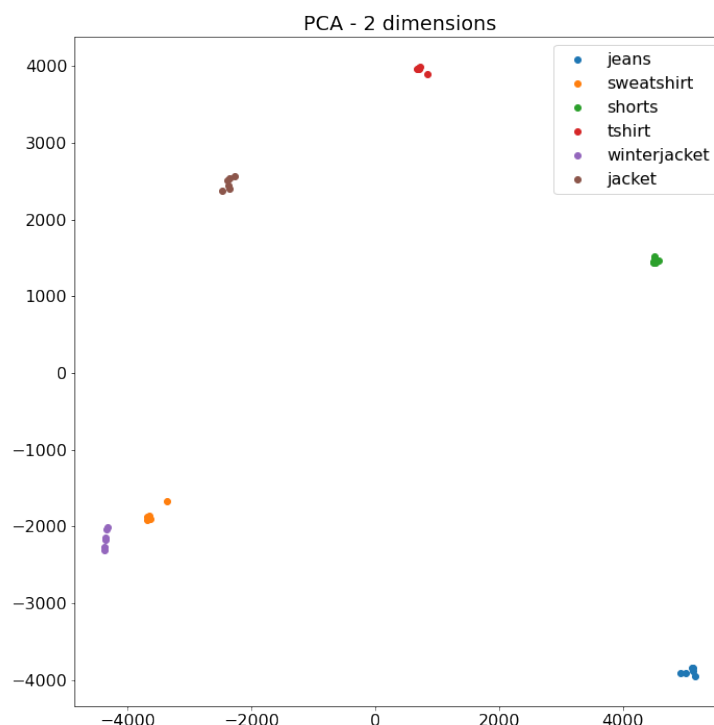
## 2.5 Redukcja do dwóch wymiarów

Na koniec, wykonano redukcję do 2 wymiarów. Zdjęcia po przekonwertowaniu (rys. 7) nie są idealne, lecz z łatwością można odróżnić poszczególne klasy od siebie. Jedyne problemem może być odróżnienie bluzy (sweatshirt) i kurtki zimowej (winterjacket). Są one najbardziej podobne (podobny odcień, długi rękaw, kaptur, zamek).



**Rys. 7:** Próbkę obserwacji po zredukowaniu do 2 wymiarów i przekształceniu z powrotem na fotografię.

Dane, po zredukowaniu wymiarowości do dwóch, można przedstawić na wykresie 2D (rys. 8).



**Rys. 8:** Wykres dwóch cech po transformacji zbioru przez PCA do 2 wymiarów

Jak widać, rzut tak skomplikowanego zbioru na dwa wymiary wygląda trywialnie. Wszystkie klasy są od siebie oddzielone. K-NN bez problemu poradziłby sobie z rozpoznaniem poszczególnych klas. Najwięcej problemów byłoby ze wspomnianymi wcześniej klasami: bluzą i kurtką zimową, lecz nawet w ich przypadku k-NN powinien sobie poradzić