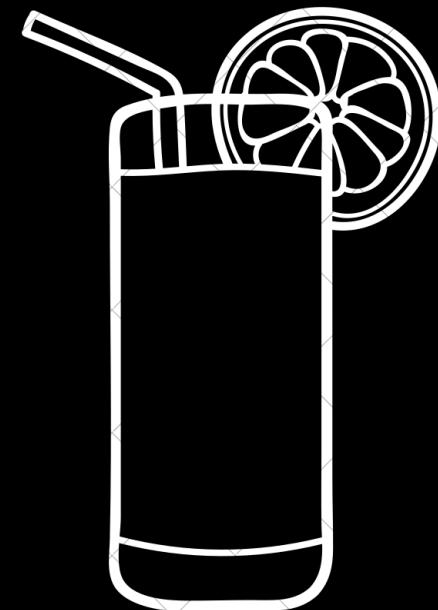
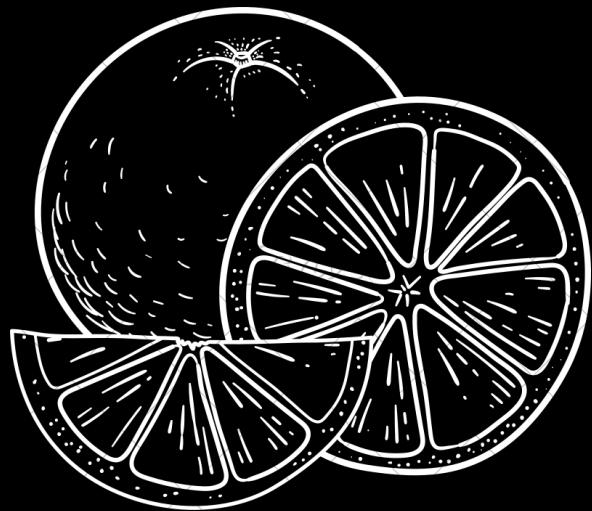


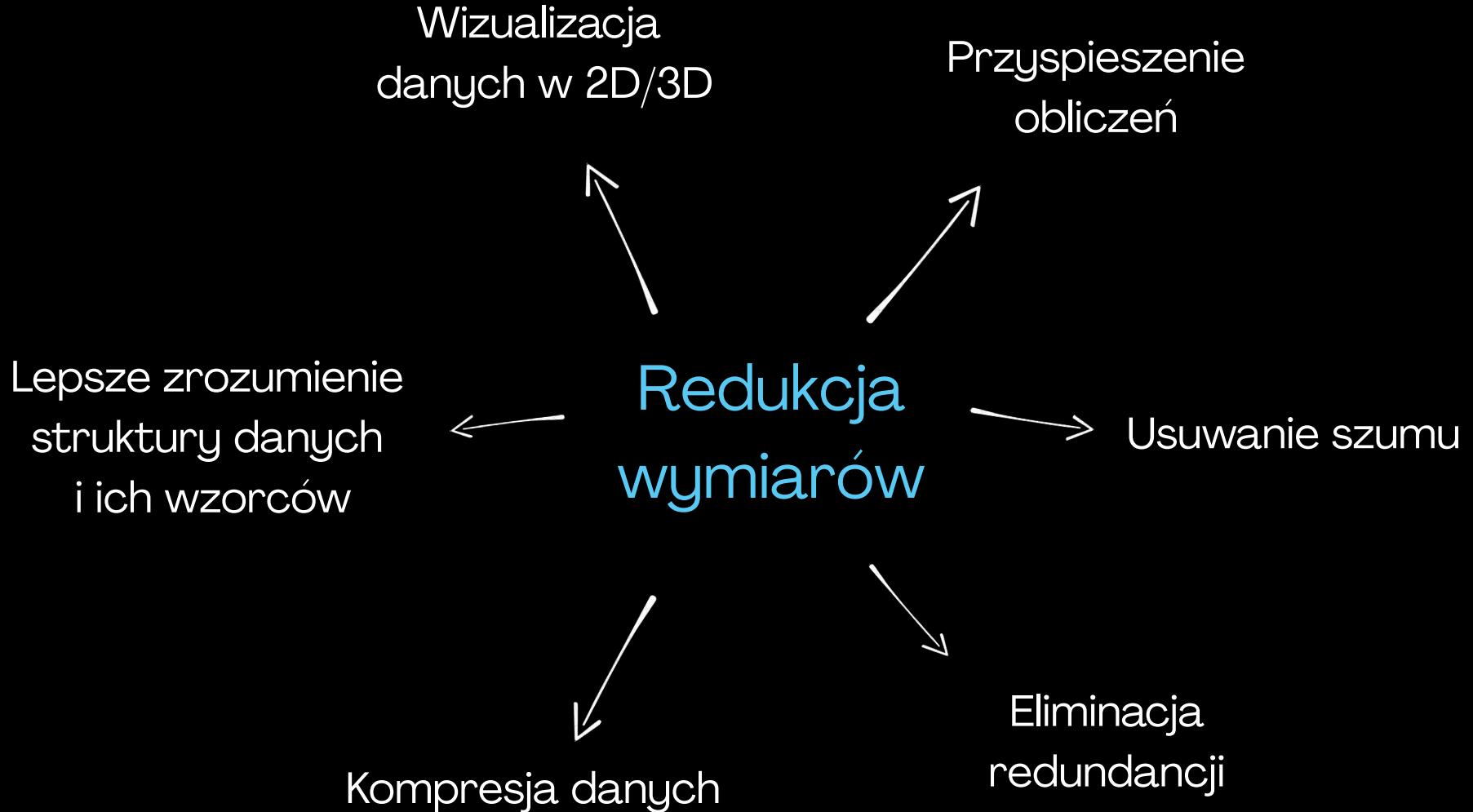
Redukcja wymiarów = wyciskarka



Po co nam redukcja wymiarów?



# Analiza danych



# Klątwa wielowymiarowości



# Klątwa wielowymiarowości

Wraz ze wzrostem liczby wymiarów:

dane stają się coraz rzadsze,

odległości między punktami tracą sens,

modele uczą się wolniej i łatwiej przeuczyć model,  
rośnie koszt obliczeniowy.

# Metody liniowe

## PCA

(Principal Component Analysis)

najpopularniejsza, przekształca dane w nowe, ortogonalne osie maksymalnej wariancji

## LDA

(Linear Discriminant Analysis)

uwzględnia przynależność do klas optymalna dla klasyfikacji

## ICA

(Independent Component Analysis)

szuka komponentów statystycznie niezależnych

# Metody nieliniowe

## t-SNE

(t-distributed Stochastic Neighbor Embedding)

świetna do wizualizacji, zachowuje lokalną strukturę danych

## UMAP

(Uniform Manifold Approximation and  
Projection)

szysze od t-SNE, zachowuje więcej  
globalnej struktury

## Isomap

oparta na geodezyjnych odległościach  
na kolektorze danych

# Metody oparte o uczenie maszynowe

## **Autoenkodery (Autoencoders)**

sieci neuronowe uczące się kompresji danych

Feature selection (np. LASSO, SelectKBest)

wybierają najważniejsze cechy bez tworzenia nowych

# Jak działa PCA?



**1**

Standaryzacja danych

**2**

Obliczenie macierzy kowariancji/korelacji

**3**

Wyznaczenie wektorów i wartości własnych

**4**

Wybór składowych głównych

**5**

Transformacja danych

# Case study

Dane z Biura Statystyk Pracy ze Stanów Zjednoczonych.  
Przedstawiają tygodniowe zarobki w 2020 roku z podziałem na stany.

# Macierz korelacji

	Transformowana liczba pracowników	Transformowana mediana zarobków	Transformowana liczba kobiet pracujących	Mediana zarobków kobiet	Transformowana ilość mężczyzn pracujących	Mediana zarobków mężczyzn
Transformowana liczba pracowników	1.000000	0.287842	0.999026	0.300256	0.999373	0.262250
Transformowana mediana zarobków	0.287842	1.000000	0.284447	0.969466	0.290151	0.969635
Transformowana liczba kobiet pracujących	0.999026	0.284447	1.000000	0.299753	0.996842	0.260402
Mediana zarobków kobiet	0.300256	0.969466	0.299753	1.000000	0.300320	0.901497
Transformowana ilość mężczyzn pracujących	0.999373	0.290151	0.996842	0.300320	1.000000	0.263151
Mediana zarobków mężczyzn	0.262250	0.969635	0.260402	0.901497	0.263151	1.000000

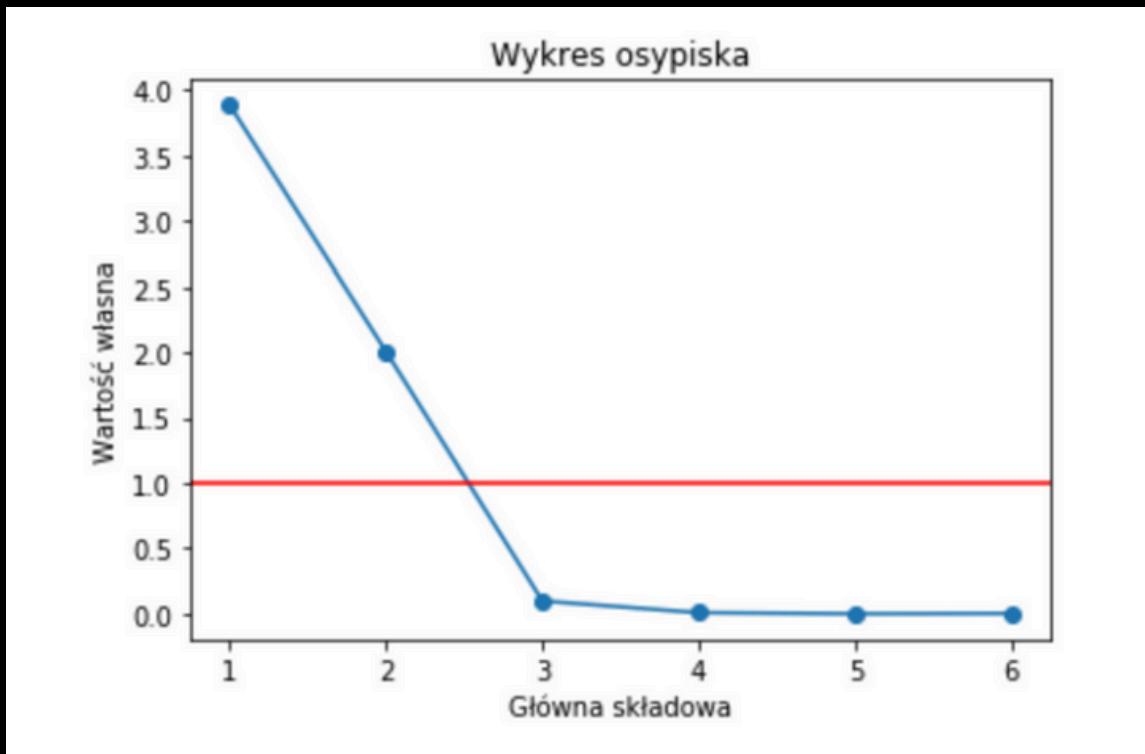
# Wektory i wartości własne

	Wektory własne					
Transformowana liczba pracowników	-0.419575	-0.396923	0.010384	0.019246	-0.814499	0.050215
Transformowana mediana zarobków	-0.403576	0.425013	-0.003109	0.735537	-0.002831	-0.339790
Transformowana liczba kobiet pracujących	-0.418716	-0.397564	0.000062	-0.305947	0.361194	-0.665248
Mediana zarobków kobiet	-0.398912	0.406232	-0.709910	-0.373295	0.000774	0.180341
Transformowana ilość mężczyzn pracujących	-0.419611	-0.395872	0.016430	0.284961	0.454001	0.616135
Mediana zarobków mężczyzn	-0.388015	0.426637	0.704017	-0.380078	0.002155	0.165289
Wartości własne	3.884728	2.006370	0.098145	0.008493	0.000001	0.002262
% Wyjaśnianej wariancji	64,75 %	33,43 %	1,64 %	0,14 %	0,00025 %	0,04 %

# Macierz ładunków czynnikowych

	PC1	PC2	PC3	PC4	PC5	PC6
<b>Transformowana liczba pracowników</b>	-0.826970	-0.795437	-0.825277	-7.862446e-01	-0.827041	-0.764767
<b>Transformowana mediana zarobków</b>	-0.562228	0.602016	-0.563135	5.754130e-01	-0.560739	0.604316
<b>Transformowana liczba kobiet pracujących</b>	0.003253	-0.000974	0.000020	-2.224010e-01	0.005147	0.220555
<b>Mediana zarobków kobiet</b>	0.001774	0.067786	-0.028195	-3.440211e-02	0.026261	-0.035027
<b>Transformowana ilość mężczyzn pracujących</b>	-0.000991	-0.000003	0.000439	9.412193e-07	0.000552	0.000003
<b>Mediana zarobków mężczyzn</b>	0.002388	-0.016162	-0.031641	8.577637e-03	0.029305	0.007862

# Wybór ilości głównych składowych



kryterium osypiska (1966)

kryterium Kaisera (1960)

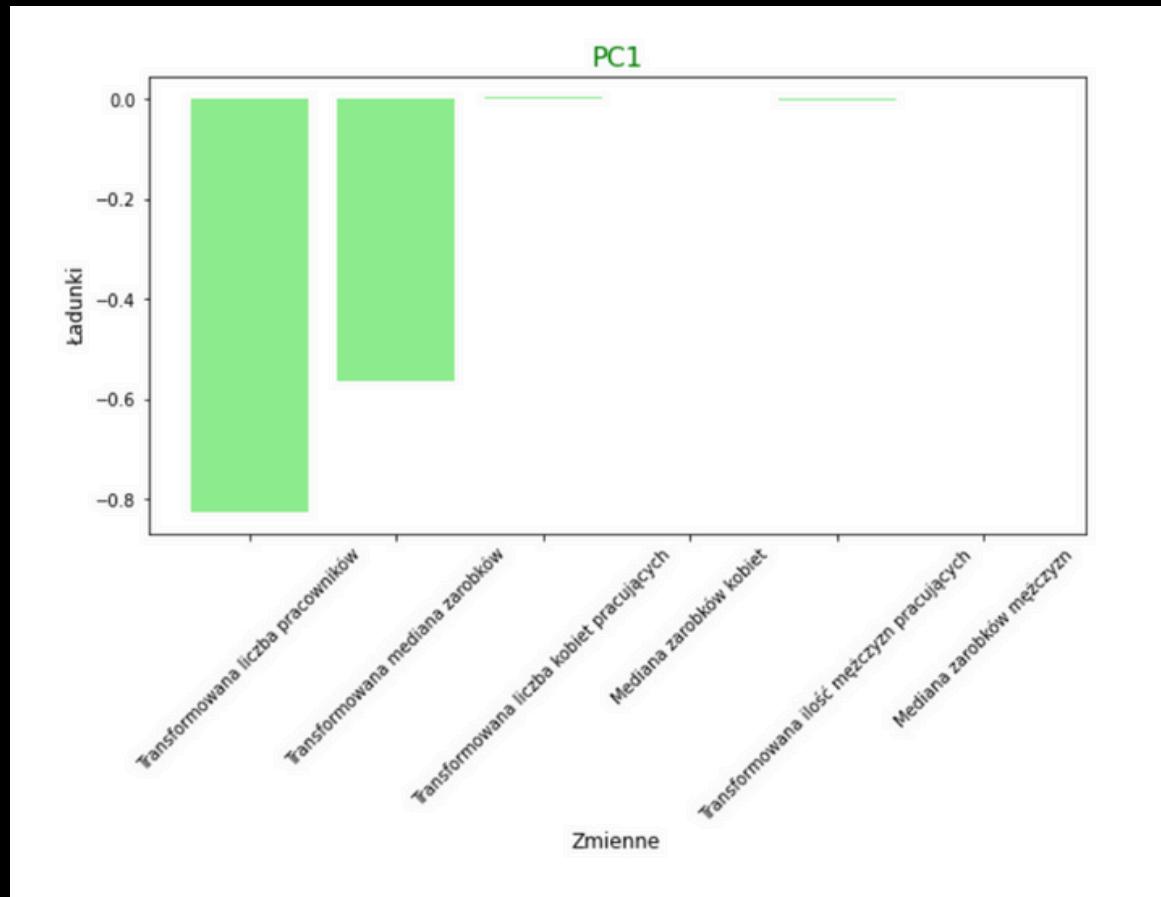
kryterium minimalnego zasobu wyjaśnianej zmienności (>70%)

$$64.75 \% + 33.43 \% = 98.18 \%$$

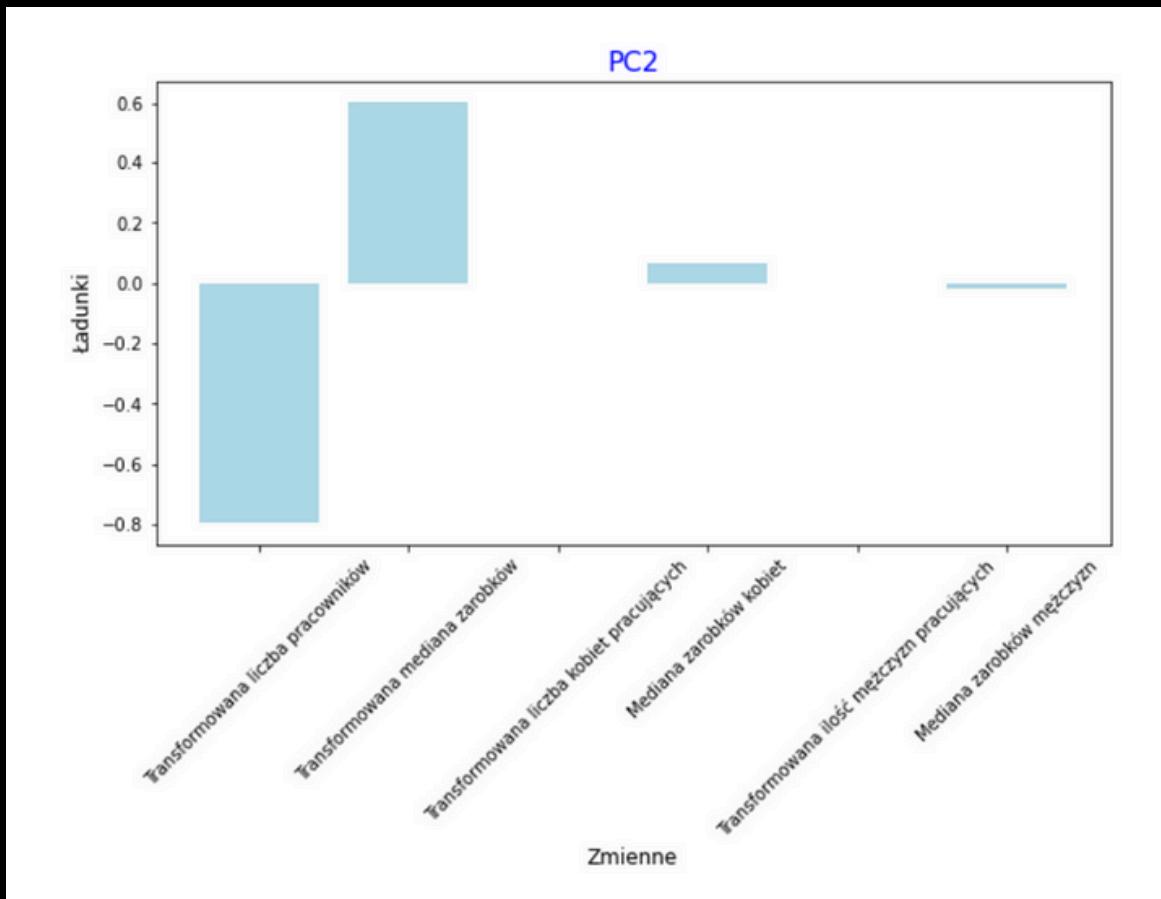
# Wybór ilości głównych składowych

	PC1	PC2
<b>Transformowana liczba pracowników</b>	-0.826970	-0.795437
<b>Transformowana mediana zarobków</b>	-0.562228	0.602016
<b>Transformowana liczba kobiet pracujących</b>	0.003253	-0.000974
<b>Mediana zarobków kobiet</b>	0.001774	0.067786
<b>Transformowana ilość mężczyzn pracujących</b>	-0.000991	-0.000003
<b>Mediana zarobków mężczyzn</b>	0.002388	-0.016162

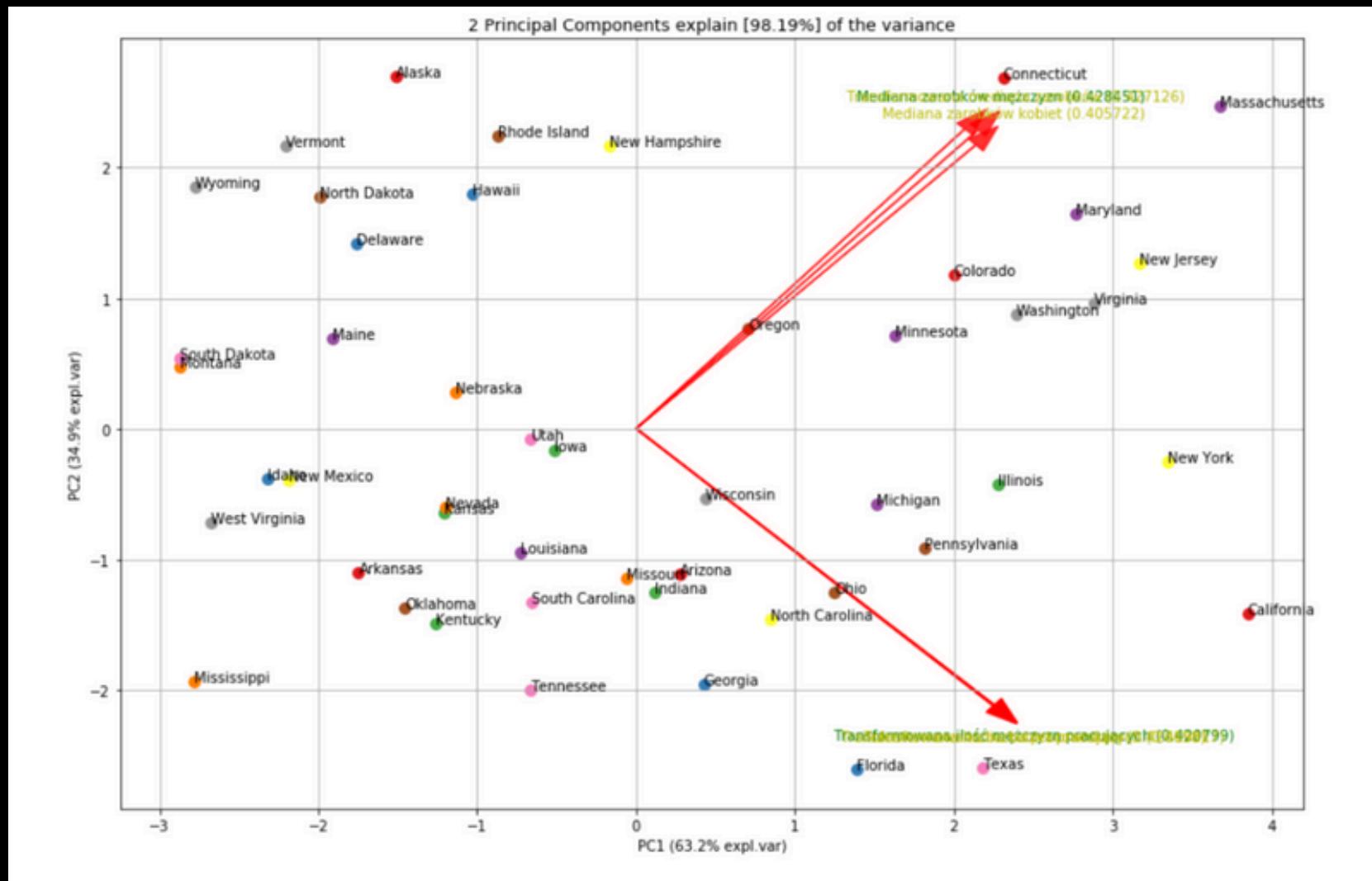
# Wykres ładunków czynnikowych



# Wykres ładunków czynnikowych



# Biplot



# Wnioski

PC1 silnie zależy od zarobków (kobiet i mężczyzn) – rozróżnia stany bogatsze i biedniejsze.

PC2 związany z udziałem pracujących mężczyzn.

Connecticut, Massachusetts, New Jersey – wysokie zarobki.

Texas, Florida, Georgia – więcej pracujących mężczyzn, ale niższe zarobki.



PCA ujawnia naturalne klastry stanów wg cech społeczno-ekonomicznych.

# Jak działa t-SNE?



# t-Distributed Stochastic Neighbor Embedding

stochastyczna metoda porządkowania sąsiadów  
w oparciu o rozkład t

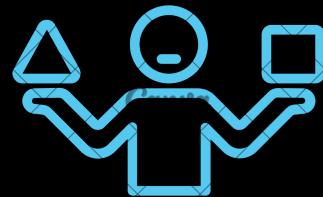




zachowuje lokalne struktury danych,  
ale niekoniecznie globalne



świetne do wizualizacji, ale nie nadaje się do predykcji



wyniki mogą się różnić przy każdym uruchomieniu  
(chyba że ustawią się random\_state)



Obliczenie podobieństw między punktami w przestrzeni wysokowymiarowej



Losowa inicjalizacja punktów w przestrzeni 2D lub 3D



Obliczenie podobieństw między punktami w niskim wymiarze



Minimalizacja różnicy między tymi dwoma rozkładami podobieństw



Przemieszczanie punktów w 2D, aż rozkłady będą jak najbardziej zbliżone

# Perplexity

Mała wartość (5–30):

t-SNE skupia się bardziej na lokalnej strukturze danych.

Może uwydatnić małe klastry.

Większe ryzyko szumu i przetrenowania.

Większa wartość (50–100):

t-SNE patrzy szerzej, uwzględnia więcej sąsiadów.

Zachowuje więcej globalnej struktury.

Może spłaszczyć lokalne różnice.

# PCA vs t-SNE

szybki, liniowy, deterministyczny

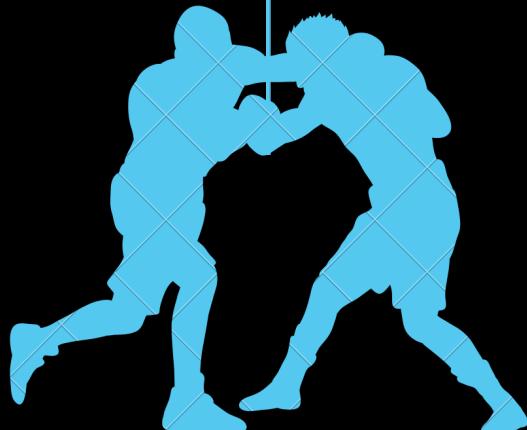
zachowuje globalną strukturę  
(maksymalizuje wariancję)

szybka redukcja wymiarów

wolniejszy, nieliniowy, probabilistyczny

lepiej oddaje lokalne podobieństwa

wizualizacja złożonych zbiorów danych



# Jak działa LDA?



# Liniowa analiza dyskryminacyjna

metoda uczenia nadzorowanego, która znajduje takie kombinacje cech, które najlepiej oddzielają od siebie różne klasy. Jej celem jest zmniejszenie liczby wymiarów, zachowując przy tym maksymalną ilość informacji potrzebną do rozróżnienia kategorii.

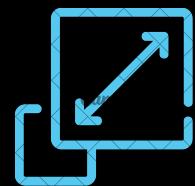
Uzyskane nowe cechy mogą być wykorzystane do budowy klasyfikatora lub jako wstępny krok w analizie danych.



super do klasyfikacji i wizualizacji,  
jeśli mamy dane z etykietami



zakłada, że dane w klasach  
są normalnie rozłożone i mają równą kowariancję



w odróżnieniu od PCA,  
LDA maksymalizuje separację klas, a nie wariancję ogólną

# CEL



Znalezienie kombinacji liniowych cech, które najlepiej odróżniają klasy w danych.

1

## Obliczenie średnich klasowych

*Dla każdej klasy obliczamy wektor średnich wartości cech*

2

## Obliczenie macierzy rozrzutu wewnętrzklasowego

*Reprezentuje wariancję cech w obrębie każdej klasy*

3

## Obliczenie macierzy rozrzutu międzyklasowego

*Reprezentuje wariancję cech między średnimi wartościami cech różnych klas.*

4

## Rozwiążanie problemu wartości własnych

*Pozwala znaleźć kierunki w przestrzeni cech, które maksymalizują separację między klasami*

5

## Wybór głównych kierunków dyskryminacji

*Wybieramy  $k-1$  wektorów własnych odpowiadających największym wartościom własnym*

6

## Transformacja danych

*Rzutujemy oryginalne dane na wybrane kierunki, uzyskując nową przestrzeń o mniejszej liczbie wymiarów*

## Rozpoznawanie płci na podstawie wzrostu i masy ciała

Osoba	Wzrost (cm)	Masa (kg)	Płeć
A	160	50	Kobieta (0)
B	165	55	Kobieta (0)
C	170	54	Kobieta (0)
D	180	75	Mężczyzna (1)
E	175	70	Mężczyzna (1)
F	185	85	Mężczyzna (1)

1

## Obliczenie średnich klasowych

Dla każdej klasy obliczamy wektor średnich wartości cech

**Kobiety (klasa 0):**

$$\mu_0 = \left[ \frac{160 + 165 + 170}{3}, \frac{50 + 55 + 54}{3} \right] = [165, 53]$$

**Mężczyźni (klasa 1):**

$$\mu_1 = \left[ \frac{180 + 175 + 185}{3}, \frac{75 + 70 + 85}{3} \right] = [180, 76.7]$$

**Średnia globalna:**

$$\mu = \left[ \frac{160 + 165 + 170 + 180 + 175 + 185}{6}, \frac{50 + 55 + 54 + 75 + 70 + 85}{6} \right] = [172.5, 66.5]$$

## Obliczenie macierzy rozrzutu wewnętrzklasowego

Reprezentuje wariancję cech w obrębie każdej klasy

Dla klasy 0 (kobiety):

Próbka	$x_i - \mu_0$	$(x_i - \mu_0)(x_i - \mu_0)^T$
A	[-5, -3]	$\begin{bmatrix} 25 & 15 \\ 15 & 9 \end{bmatrix}$
B	[0, 2]	$\begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$
C	[5, 1]	$\begin{bmatrix} 25 & 5 \\ 5 & 1 \end{bmatrix}$

$$S_{W0} = \sum = \begin{bmatrix} 50 & 20 \\ 20 & 14 \end{bmatrix}$$

# Obliczenie macierzy rozrzutu wewnętrzklasowego

*Reprezentuje wariancję cech w obrębie każdej klasy*

Dla klasy 1 (mężczyźni):

Próbka	$x_i - \mu_1$	$(x_i - \mu_1)(x_i - \mu_1)^T$
D	[0, -1.7]	$\begin{bmatrix} 0 & 0 \\ 0 & 2.89 \end{bmatrix}$
E	[-5, -6.7]	$\begin{bmatrix} 25 & 33.5 \\ 33.5 & 44.89 \end{bmatrix}$
F	[5, 8.3]	$\begin{bmatrix} 25 & 41.5 \\ 41.5 & 68.89 \end{bmatrix}$

$$S_{W1} = \sum = \begin{bmatrix} 50 & 75 \\ 75 & 116.67 \end{bmatrix}$$

Całkowita macierz  $S_W$ :

$$S_W = S_{W0} + S_{W1} = \begin{bmatrix} 100 & 95 \\ 95 & 130.67 \end{bmatrix}$$

3

## Obliczenie macierzy rozrzutu międzyklasowego

Reprezentuje wariancję cech między średnimi wartościami cech różnych klas.

$$S_B = \sum n_k(\mu_k - \mu)(\mu_k - \mu)^T$$

Dla klasy 0:

$$\mu_0 - \mu = [165 - 172.5, 53 - 66.5] = [-7.5, -13.5] \Rightarrow 3 \cdot [-7.5, -13.5]^T [-7.5, -13.5] = 3 \cdot \begin{bmatrix} 56.25 & 101.25 \\ 101.25 & 182.25 \end{bmatrix}$$

Dla klasy 1:

$$\mu_1 - \mu = [180 - 172.5, 76.7 - 66.5] = [7.5, 10.2] \Rightarrow 3 \cdot \begin{bmatrix} 56.25 & 76.5 \\ 76.5 & 104.04 \end{bmatrix}$$

**Suma:**

$$S_B = \begin{bmatrix} 337.5 & 532.5 \\ 532.5 & 858.87 \end{bmatrix}$$

4

## Rozwiązańe problemu wartości własnych

*Pozwala znaleźć kierunki w przestrzeni cech, które maksymalizują separację między klasami*

Obliczamy:

$$S_W^{-1} S_B$$

Znajdujemy największą wartość własną i jej wektor,  
to będzie kierunek LDA  
(przy małej liczbie cech: korzystamy z numpy  
lub rysujemy geometrycznie).

5

## Wybór głównych kierunków dyskryminacji

*Wybieramy  $k-1$  wektorów własnych odpowiadających największym wartościom własnym*

$$k-1 = 2-1 = 1$$

## Transformacja danych

Rzutujemy oryginalne dane na wybrane kierunki, uzyskując nową przestrzeń o mniejszej liczbie wymiarów

Każdy punkt:

$$z_i = w^T x_i$$

daje jedną wartość liczbową = współrzędna na osi LDA (1D).

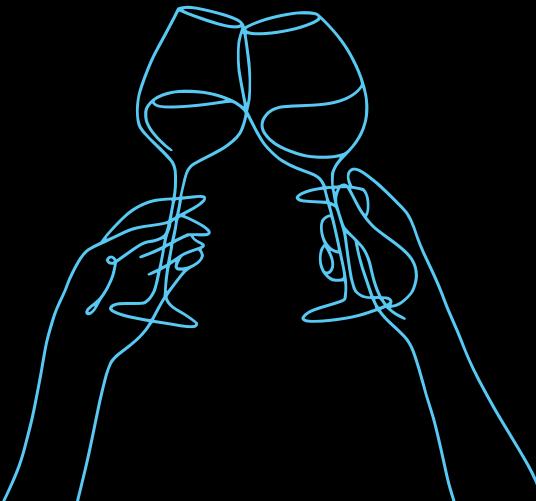
# Case study

Zbiór danych Wine z biblioteki scikit-learn.

Liczba klas: 3 (trzy różne odmiany wina)

Liczba cech: 13 (chemiczne właściwości win)

Liczba próbek: 178



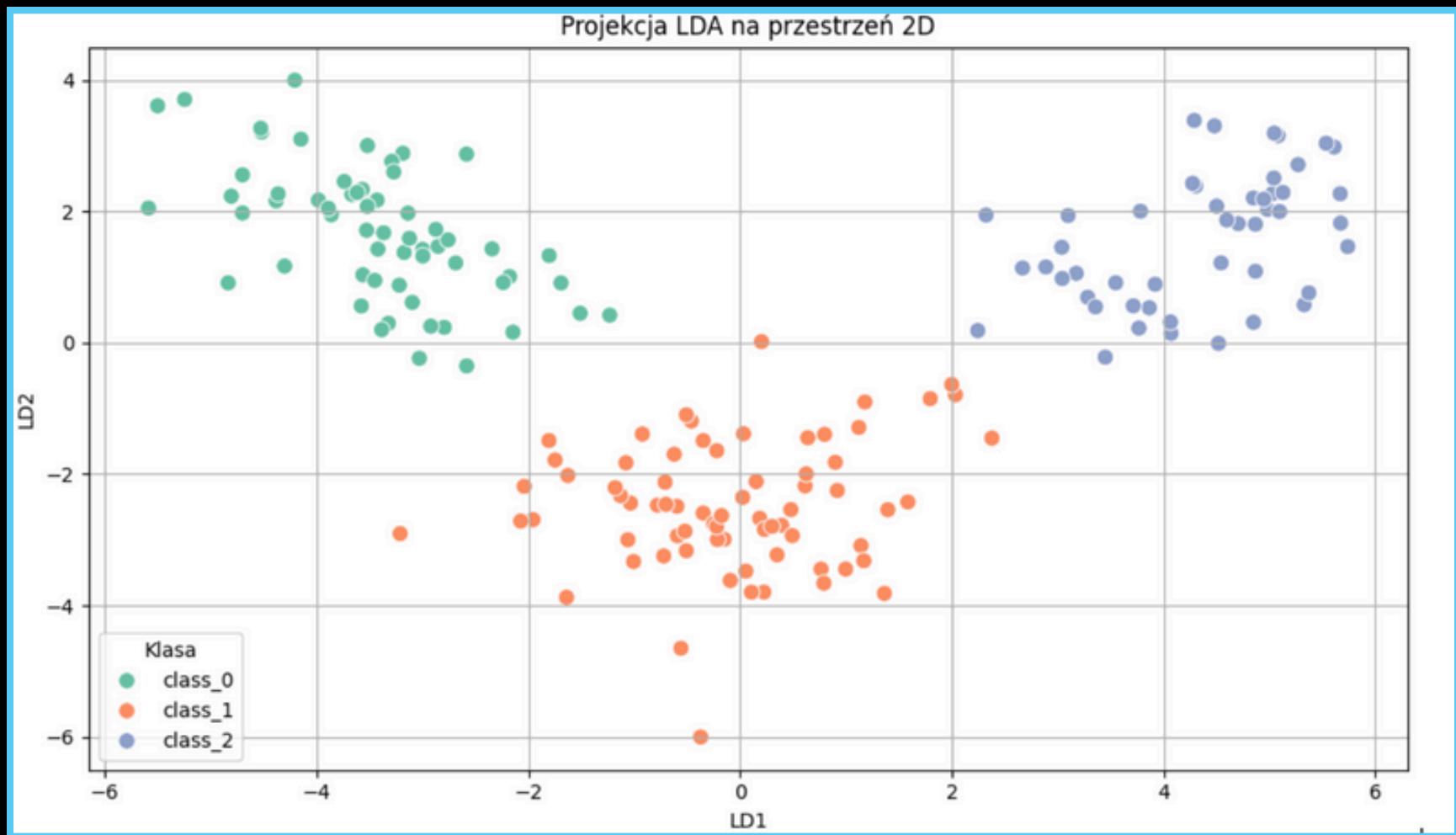
# Przegląd danych

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavanoid_phenols	proanthocyanins	color_intensity
0	14.23	1.71	2.43	15.6	127.0	2.80	3.06	0.28	2.29	5.64
1	13.20	1.78	2.14	11.2	100.0	2.65	2.76	0.26	1.28	4.38
2	13.16	2.36	2.67	18.6	101.0	2.80	3.24	0.30	2.81	5.68
3	14.37	1.95	2.50	16.8	113.0	3.85	3.49	0.24	2.18	7.80
4	13.24	2.59	2.87	21.0	118.0	2.80	2.69	0.39	1.82	4.32
...	...	...	...	...	...	...	...	...	...	...
173	13.71	5.65	2.45	20.5	95.0	1.68	0.61	0.52	1.06	7.70
174	13.40	3.91	2.48	23.0	102.0	1.80	0.75	0.43	1.41	7.30
175	13.27	4.28	2.26	20.0	120.0	1.59	0.69	0.43	1.35	10.20
176	13.17	2.59	2.37	20.0	120.0	1.65	0.68	0.53	1.46	9.30
177	14.13	4.10	2.74	24.5	96.0	2.05	0.76	0.56	1.35	9.20
178 rows × 14 columns										

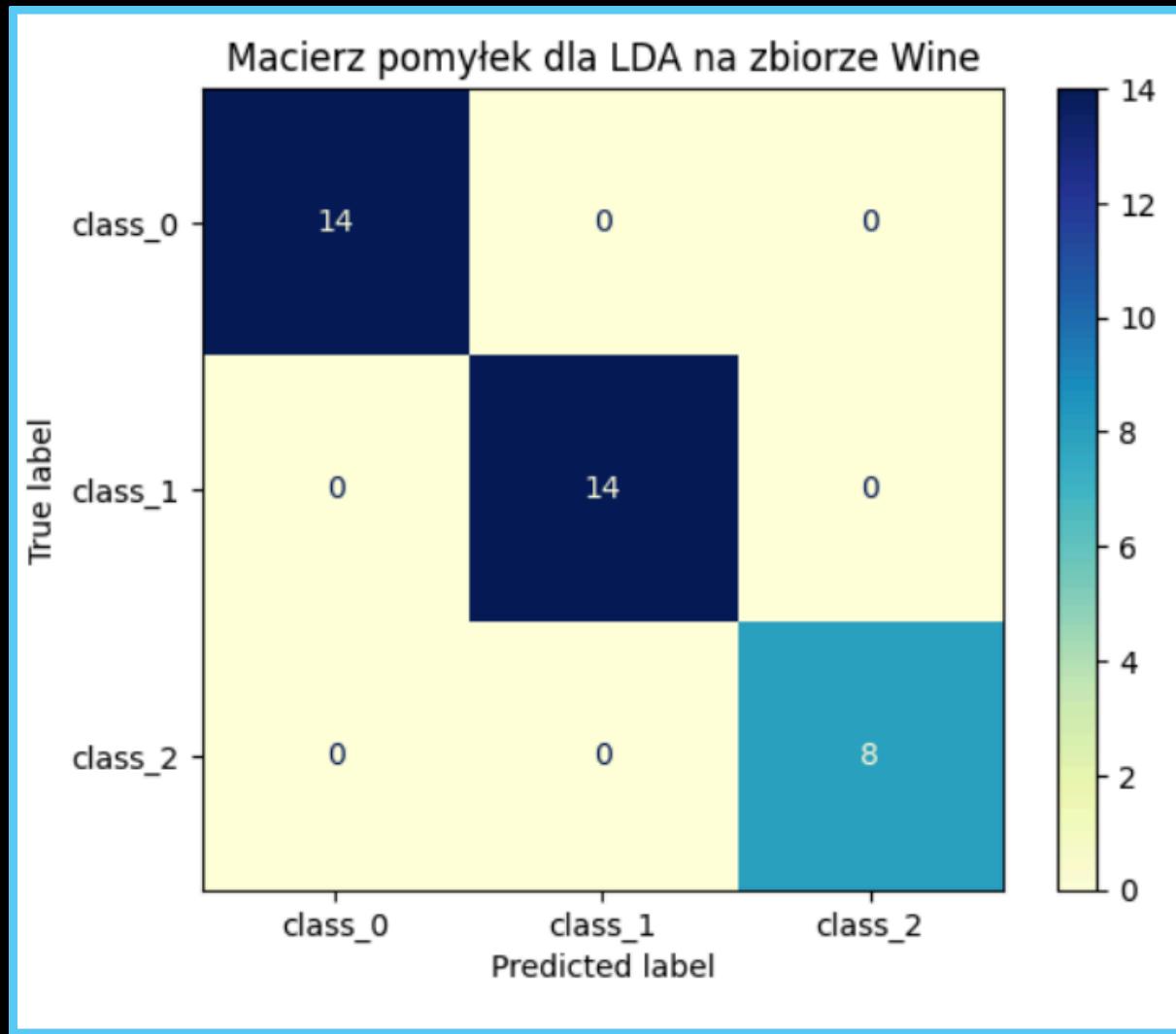
## Dane transformowane

	LD1	LD2	class	class_name
0	-4.700244	1.979138	0	class_0
1	-4.301958	1.170413	0	class_0
2	-3.420720	1.429101	0	class_0
3	-4.205754	4.002871	0	class_0
4	-1.509982	0.451224	0	class_0
...	...	...	...	...
173	4.291508	3.390332	2	class_2
174	4.503296	2.083546	2	class_2
175	5.047470	3.196231	2	class_2
176	4.276155	2.431388	2	class_2
177	5.538086	3.042057	2	class_2
178 rows × 4 columns				

# Projekcja danych



# Preidykcja



Średnia dokładność (CV): 96.63%

# LDA to nie tylko redukcja – to klasyfikator Bayesowski

LDA zakłada, że dane w każdej klasie pochodzą z  
**wielowymiarowego rozkładu normalnego.**

Zakłada też **wspólną macierz kowariancji** dla wszystkich klas.

Bazując na tym, LDA dokonuje klasyfikacji Bayesowskiej, wybierając klasę z **najwyższym prawdopodobienstwem**.

$$\frac{P(B|A) \times P(A)}{P(B)}$$

LDA to nie tylko redukcja – to klasyfikator Bayesowski



# Argumenty funkcji LDA ze scikit-learn

## **solver**

Algorytm: 'svd', 'lsqr', 'eigen'

## **shrinkage**

Regularizacja: 'auto', float, 'none'

## **n\_components**

Ile wymiarów LDA chcesz uzyskać (max = liczba klas – 1)

## **priors**

Ręczne ustawienie prawdopodobieństw klas

## **store\_covariance**

Czy przechować macierz kowariancji

## **tol**

Tolerancja numeryczna dla eigen solvera

# Jak wybrać solver?

**svd** - szybki, bez kowariancji, bez shrinkage (najczęściej używany)

**lsqr** - obsługuje shrinkage, dobry do wysokowymiarowych danych

**eigen** - jak lsqr, ale używa innej metody własnej, wolniejszy



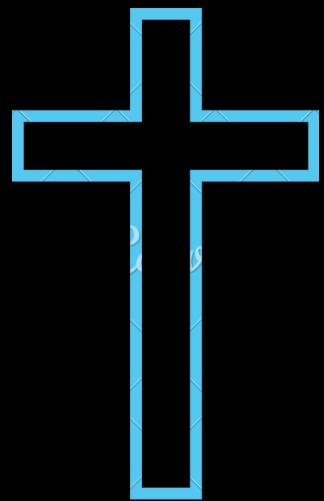
# Shrinkage

Gdy liczba cech jest wysoka względem liczby próbek, macierz kowariancji może być niestabilna. Shrinkage uśrednia ją z macierzą jednostkową.

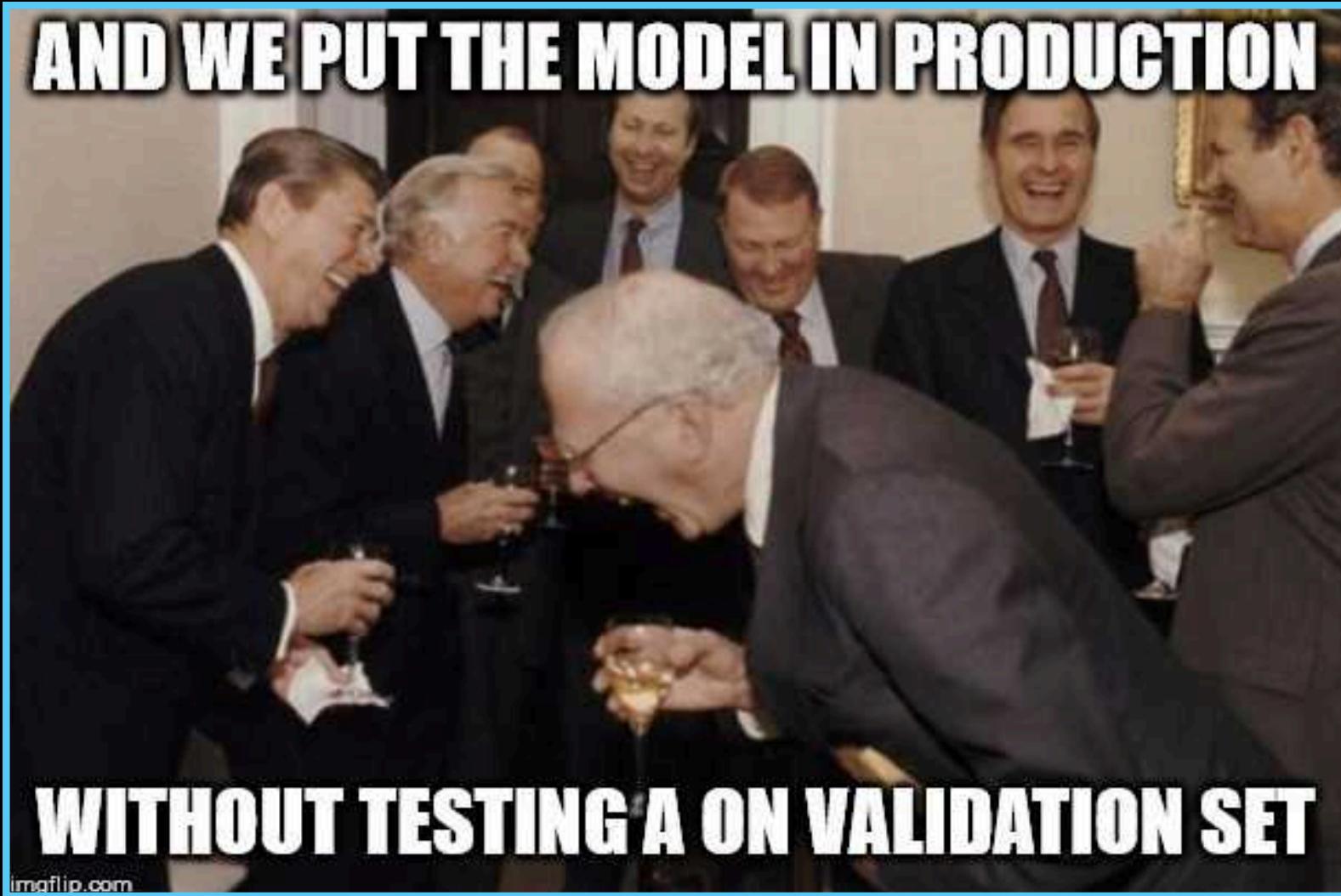
Pomaga uniknąć nadmiernego dopasowania (overfittingu).

## Walidacja krzyżowa

Pomaga oszacować, jak dobrze LDA radzi sobie z danymi nieznanymi.



**AND WE PUT THE MODEL IN PRODUCTION**



**WITHOUT TESTING A ON VALIDATION SET**

imgflip.com

# ZALETY I OGRANICZENIA LDA

-  Efektywna redukcja wymiarowości przy zachowaniu informacji o klasach.
-  Poprawa wydajności algorytmów klasyfikacyjnych.
-  Lepsza interpretowalność danych dzięki projekcji na mniejszą liczbę wymiarów.
-  Założenie liniowej separowalności klas, co może nie być spełnione w rzeczywistości.
-  Wrażliwość na obecność wartości odstających.
-  Wymaga, aby liczba próbek w każdej klasie była większa niż liczba cech.

# LDA vs PCA

metoda uczenia z nadzorem

uwzględnia etykiety klas

dąży do maksymalizacji  
separacji między klasami

metoda uczenia bez nadzoru

nie uwzględnia etykiet klas

koncentruje się na maksymalnej  
wariancji danych

