Widzenie Komputerowe - PCA, ICA

Wykład 7.

Magdalena Mazur-Milecka

Katedra Inżynierii Biomedycznej, WETI, PG

6 kwietnia 2020

PCA - Principal Component Analysis - Analiza głównych składowych

PCA - jedna ze statystycznych metod analizy czynnikowej. Zbiór danych składający się z N obserwacji, z których każda obejmuje K zmiennych, można interpretować jako chmurę N punktów w przestrzeni K-wymiarowej.

Wikipedia

PCA - Principal Component Analysis - Analiza głównych składowych

PCA - jedna ze statystycznych metod analizy czynnikowej. Zbiór danych składający się z N obserwacji, z których każda obejmuje K zmiennych, można interpretować jako chmurę N punktów w przestrzeni K-wymiarowej.
Celem PCA jest taki obrót układu współrzędnych, aby maksymalizować w pierwszej kolejności wariancję pierwszej współrzędnej, następnie wariancję drugiej współrzędnej itd..

Wikipedia

PCA - Principal Component Analysis - Analiza głównych składowych

PCA - jedna ze statystycznych metod analizy czynnikowej. Zbiór danych składający się z N obserwacji, z których każda obejmuje K zmiennych, można interpretować jako chmurę N punktów w przestrzeni K-wymiarowej.
Celem PCA jest taki obrót układu współrzędnych, aby maksymalizować w pierwszej kolejności wariancję pierwszej współrzędnej, następnie wariancję drugiej współrzędnej itd..
Tak przekształcone wartości współrzędnych nazywane są ładunkami wygenerowanych czynników (składowych głównych). W ten sposób konstruowana jest nowa przestrzeń obserwacji, w której najwięcej zmienności wyjaśniają początkowe czynniki.

Wikipedia

PCA - Principal Component Analysis - Analiza głównych składowych

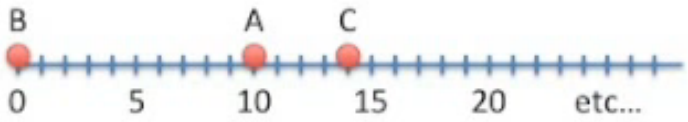
PCA - jedna ze statystycznych metod analizy czynnikowej. Zbiór danych składający się z N obserwacji, z których każda obejmuje K zmiennych, można interpretować jako chmurę N punktów w przestrzeni K-wymiarowej.
Celem PCA jest taki obrót układu współrzędnych, aby maksymalizować w pierwszej kolejności wariancję pierwszej współrzędnej, następnie wariancję drugiej współrzędnej itd..
Tak przekształcone wartości współrzędnych nazywane są ładunkami wygenerowanych czynników (składowych głównych). W ten sposób konstruowana jest nowa przestrzeń obserwacji, w której najwięcej zmienności wyjaśniają początkowe czynniki.
Ideą PCA jest ortogonalna transformacja układu badanych zmiennych X w układ nowych zmiennych nieobserwowanych Y. Zmienne Y są liniowymi kombinacjami zmiennych X i nazywane są składowymi głównymi.

Wikipedia

PCA - Definicja problemu

Załóżmy, że 3 różne obiekty (A, B i C) przyjmują 3 różne wartości dla określonego parametru (Cell 1). Można przedstawić to na 1-wymiarowej osi

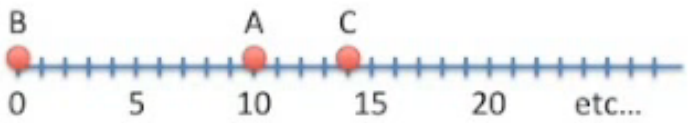
	A	B	C
Cell 1.	10	0	14



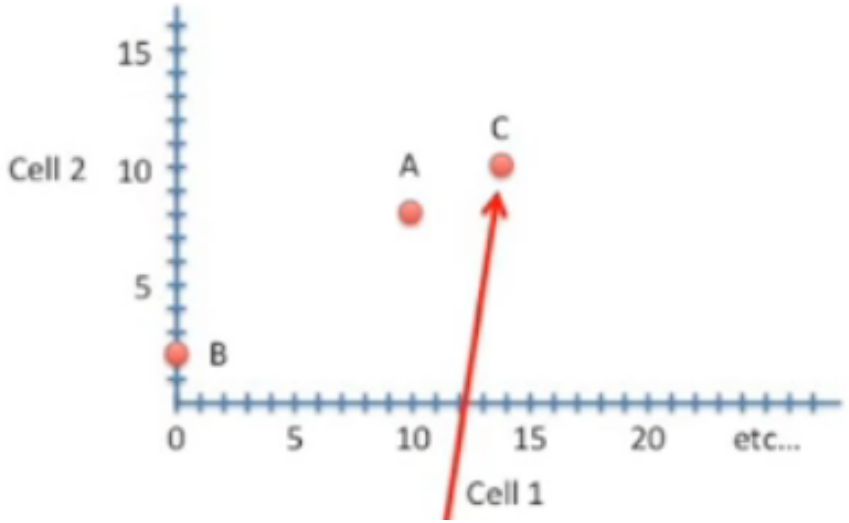
PCA - Definicja problemu

Dla dwóch parametrów (Cell 1 i Cell 2)wartości można przedstawić na osi 2D

	A	B	C
Cell 1.	10	0	14



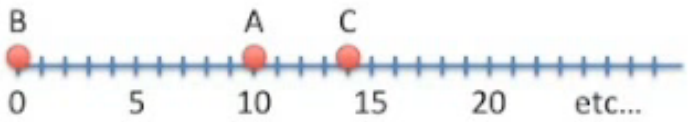
	A	B	C
Cell 1.	10	0	14
Cell 2.	8	2	10



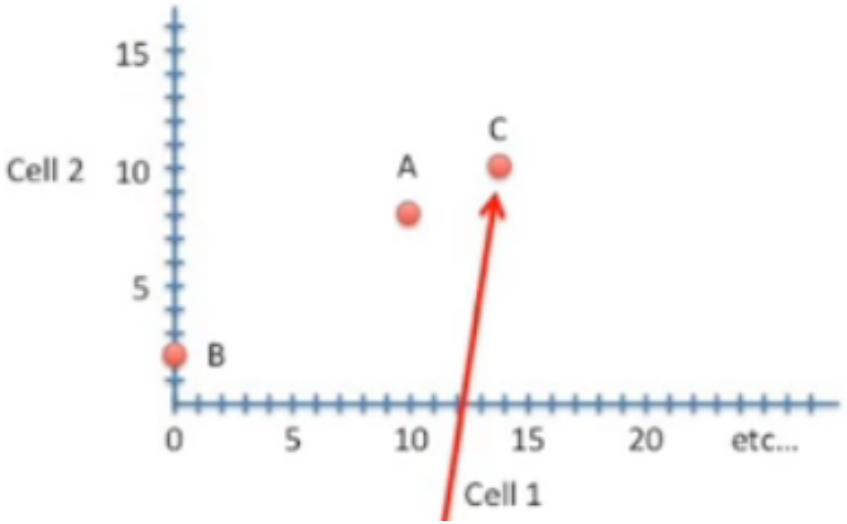
youtube.com/StatQuest

PCA - Definicja problemu

	A	B	C
Cell 1.	10	0	14



	A	B	C
Cell 1.	10	0	14
Cell 2.	8	2	10



Co w takim razie z większą ilością parametrów? Jak jesteśmy w stanie je zwizualizować?

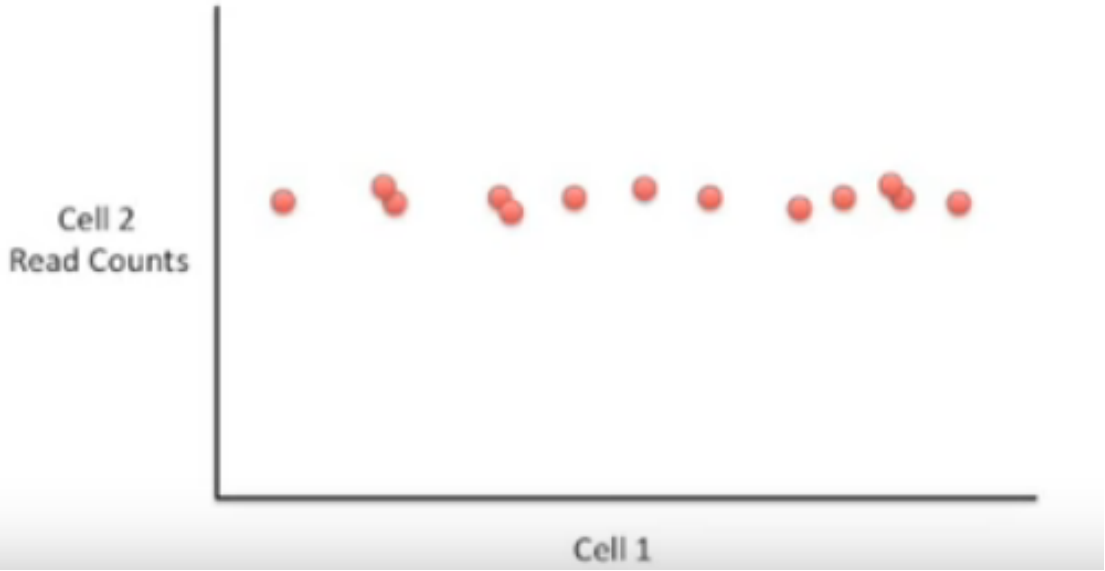
	A	B	C
Cell 1.	10	0	14
Cell 2.	8	2	10
Cell 3.	2	22	1
Cell 4.	3	4	3

youtube.com/StatQuest

Cele PCA

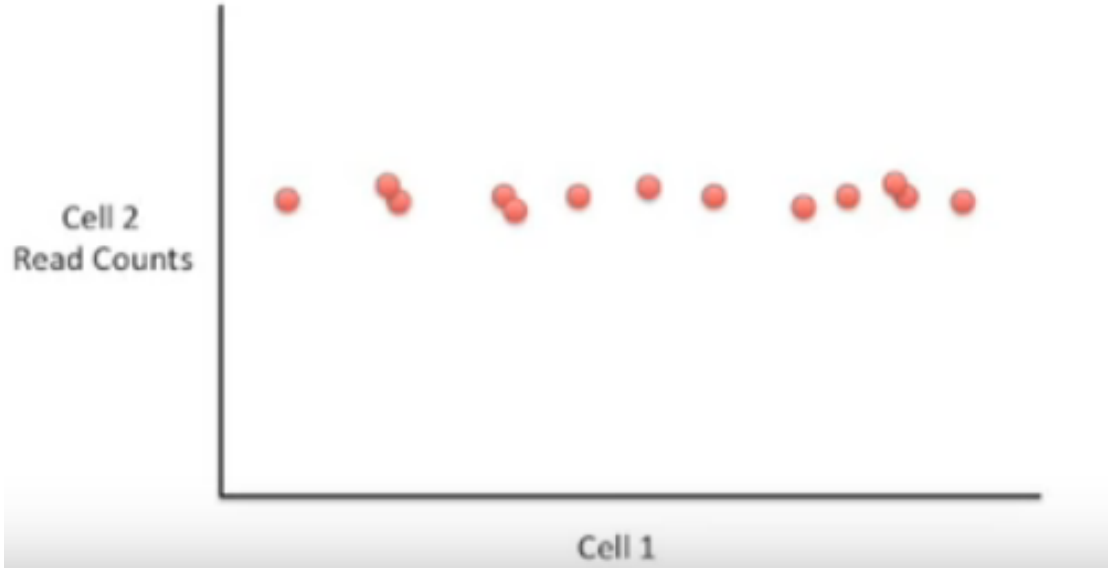
- wizualizacja pomiarów
- redukcja liczby zmiennych (wymiarowości),
- wykrywanie korelacji pomiędzy zmiennymi zbioru,
- badanie grupowania zmiennych,
- klasyfikacja obiektów w nowych przestrzeniach zdefiniowanych przez utworzone czynniki.

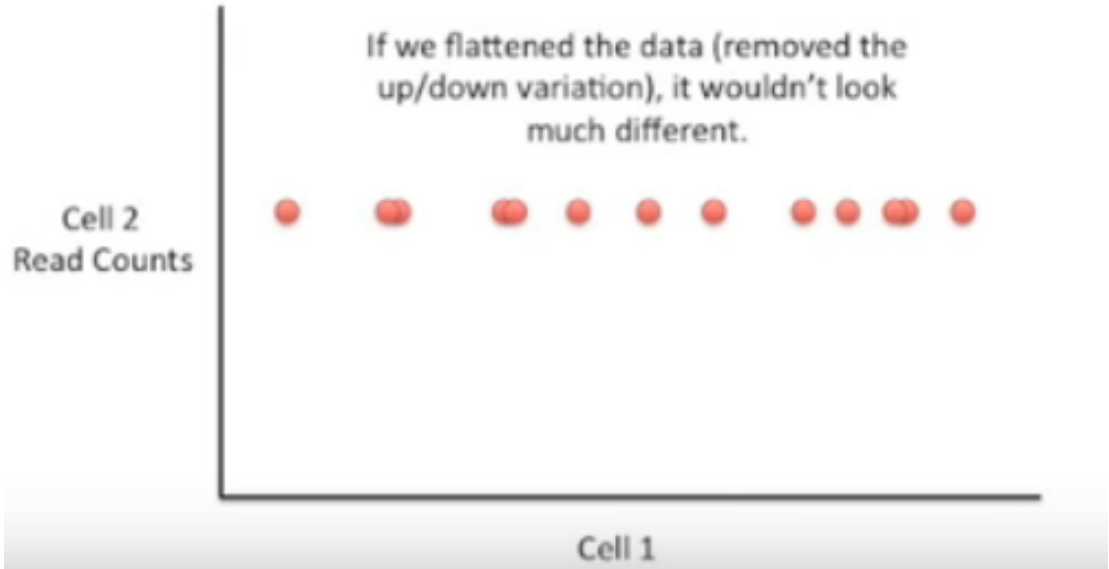
PCA - Redukcja wymiarowości



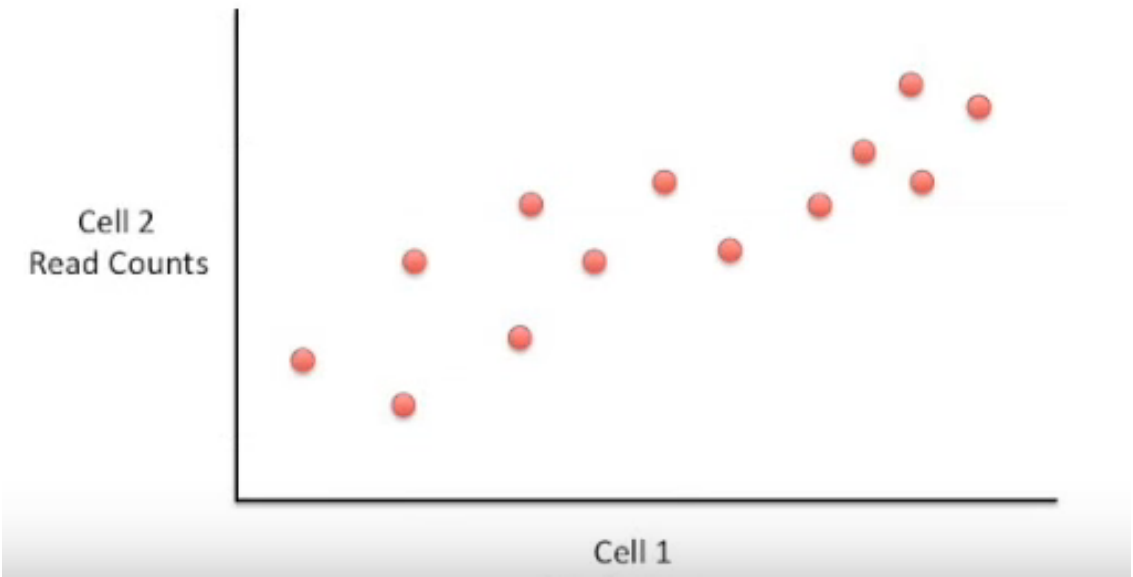
Załóżmy, że pewien parametr przyjmuje takie wartości dla różnych obiektów. Czy na jego podstawie możemy różnicować (rozpoznawać) określone obiekty?

PCA - Redukcja wymiarowości





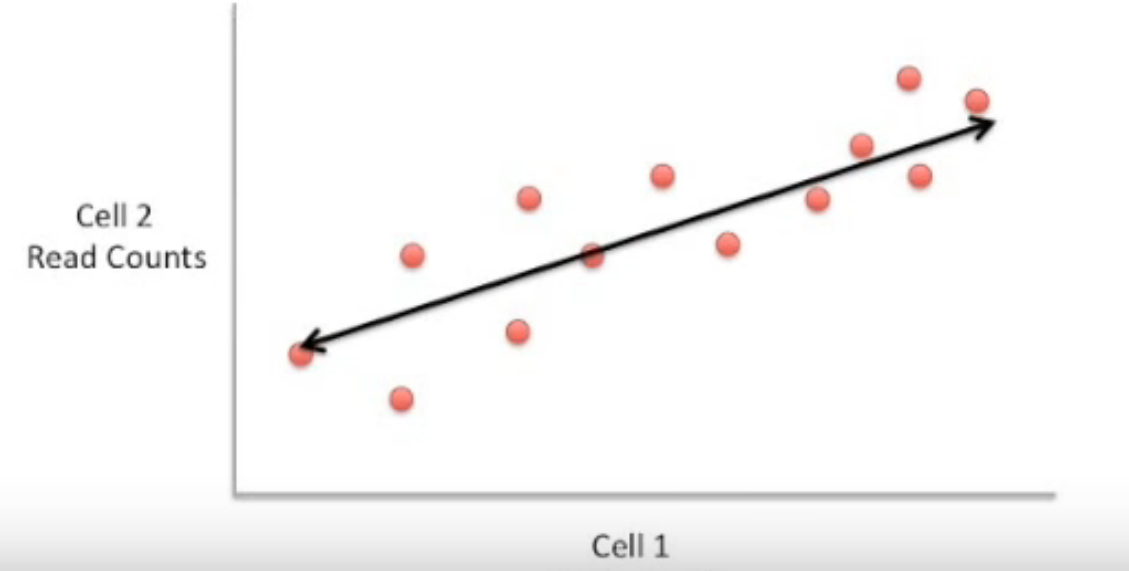
PCA - Główna idea



Wniosek

Parametry o małej wariancji nie różnicują obiektów zbyt dobrze.
Zależy nam na znalezieniu parametrów, które mają dużą wariancję.

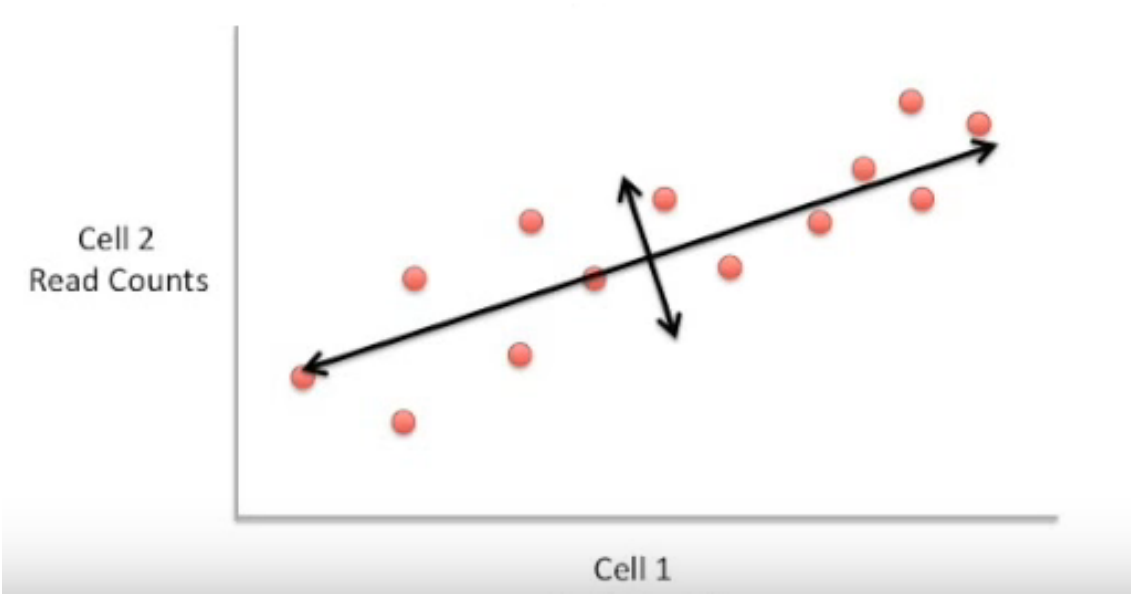
PCA - Główna idea



Parametry o największej wariancji nie zawsze są zgodne z układem współrzędnych pomiarów.

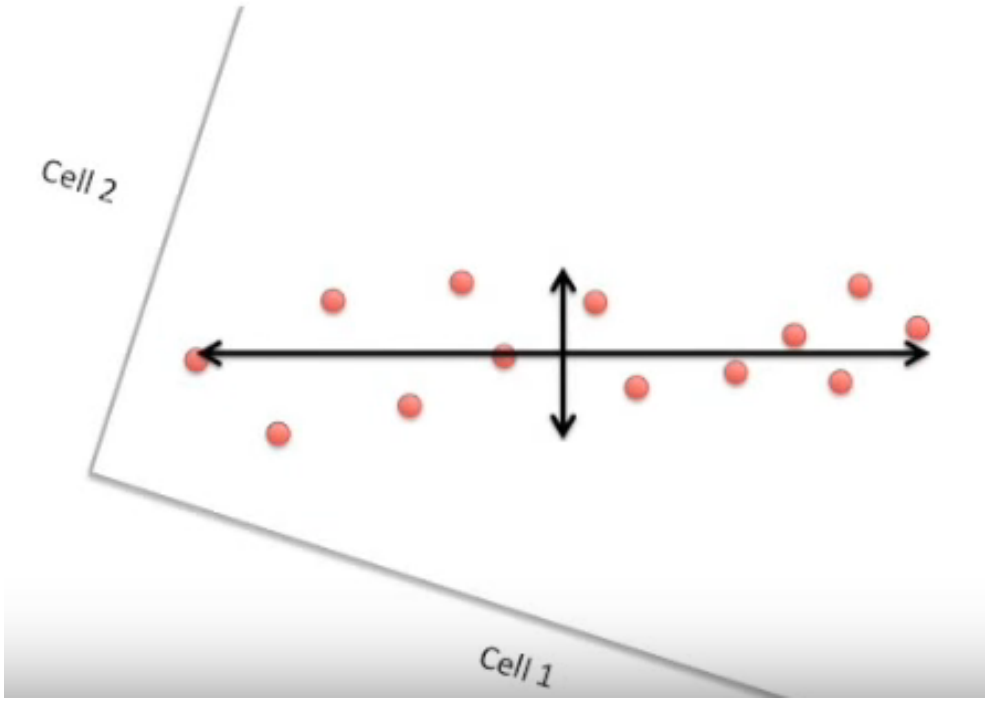
Dlatego tworzymy nowy układ współrzędnych najlepiej odpowiadający pomiarom.

PCA - Główna idea



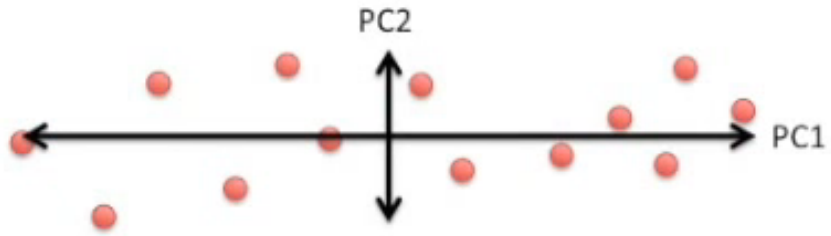
Dla nowej osi zgodnej z największa wartością wariancji znajdujemy oś ortogonalną o kolejnej największej wariancji pomiarów...

PCA - Główna idea



... i zgodnie z nimi obracamy układ współrzędnych

PCA - Główna idea



Nowy parametr (oś) o największej wariancji jest nazywany PC1, kolejne przyjmują kolejne numery (PC2, PC3 itp.

PCA - Cechy

- Nie wszystkie wymiary (badane cechy) są znaczące,

PCA - Cechy

- Nie wszystkie wymiary (badane cechy) są znaczące,
- PCA tworzy nowy układ współrzędnych (osi PC1, PC2 itd.), posortowany względem "ważności" cech,

PCA - Cechy

- Nie wszystkie wymiary (badane cechy) są znaczące,
- PCA tworzy nowy układ współrzędnych (osi PC1, PC2 itd.), posortowany względem "ważności" cech,
- PC1 jest w kierunku największej zmienności (wariancji) danych,

PCA - Cechy

- Nie wszystkie wymiary (badane cechy) są znaczące,
- PCA tworzy nowy układ współrzędnych (osi PC1, PC2 itd.), posortowany względem "ważności" cech,
- PC1 jest w kierunku największej zmienności (wariancji) danych,
- PC2 jest w kierunku 2. największej zmienności danych,

PCA - Cechy

- Nie wszystkie wymiary (badane cechy) są znaczące,
- PCA tworzy nowy układ współrzędnych (osi PC1, PC2 itd.), posortowany względem "ważności" cech,
- PC1 jest w kierunku największej zmienności (wariancji) danych,
- PC2 jest w kierunku 2. największej zmienności danych,
- PC3 .....

PCA - Cechy

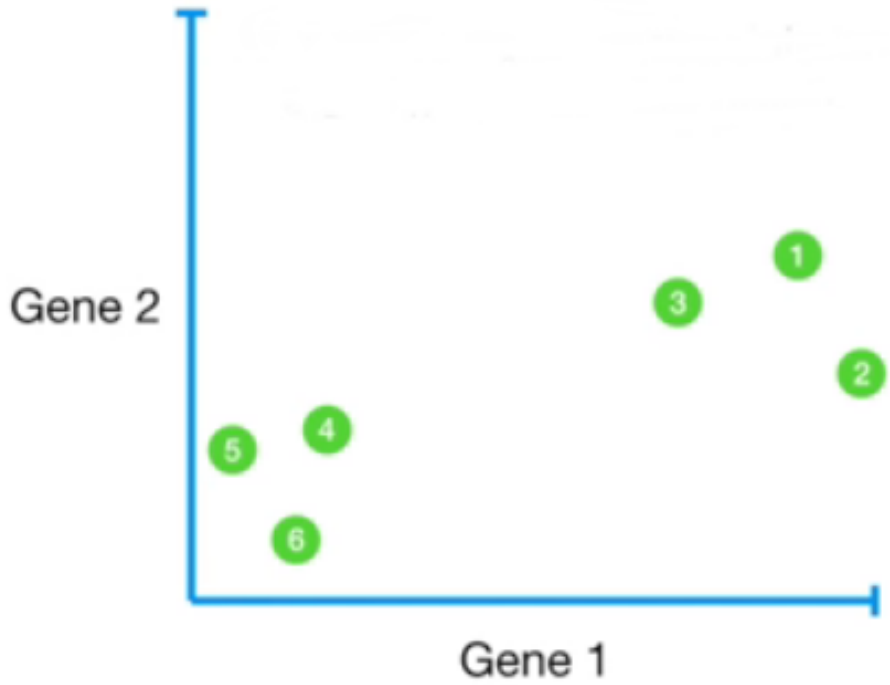
- Nie wszystkie wymiary (badane cechy) są znaczące,
- PCA tworzy nowy układ współrzędnych (osi PC1, PC2 itd.), posortowany względem "ważności" cech,
- PC1 jest w kierunku największej zmienności (wariancji) danych,
- PC2 jest w kierunku 2. największej zmienności danych,
- PC3 .....
- Składowe główne są liniową kombinacją badanych cech,

PCA - Cechy

- Nie wszystkie wymiary (badane cechy) są znaczące,
- PCA tworzy nowy układ współrzędnych (osi PC1, PC2 itd.), posortowany względem "ważności" cech,
- PC1 jest w kierunku największej zmienności (wariancji) danych,
- PC2 jest w kierunku 2. największej zmienności danych,
- PC3 .....
- Składowe główne są liniową kombinacją badanych cech,
- Ilość wymiarów przed i po PCA jest jednakowa, jednak ostatnie wymiary PC wnoszą najmniej informacji i można je pominąć.

PCA - Krok po kroku

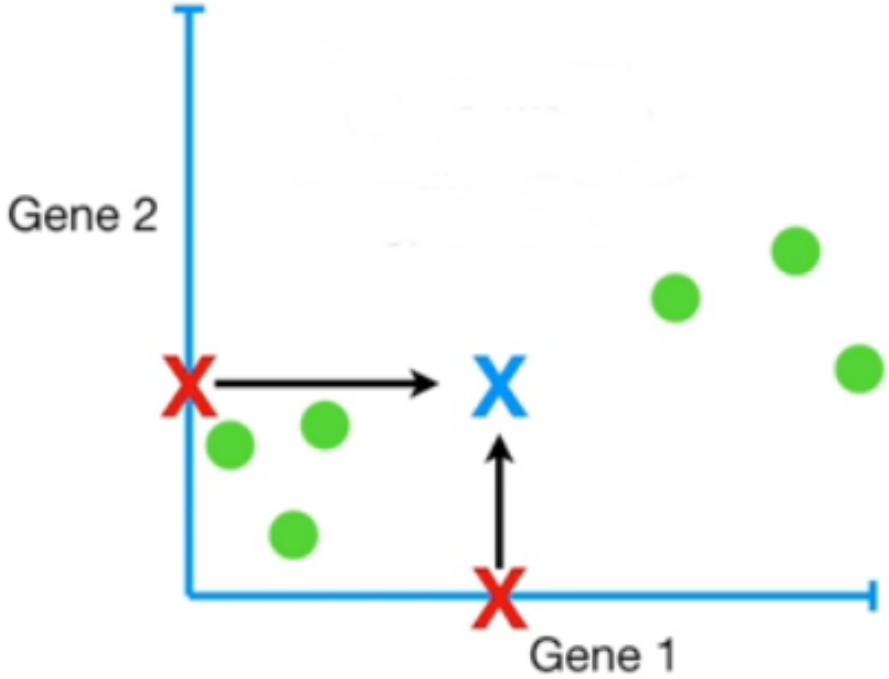
	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	2	1
Gene 2	6	4	5	3	2.8	1



youtube.com/StatQuest

PCA - Krok po kroku

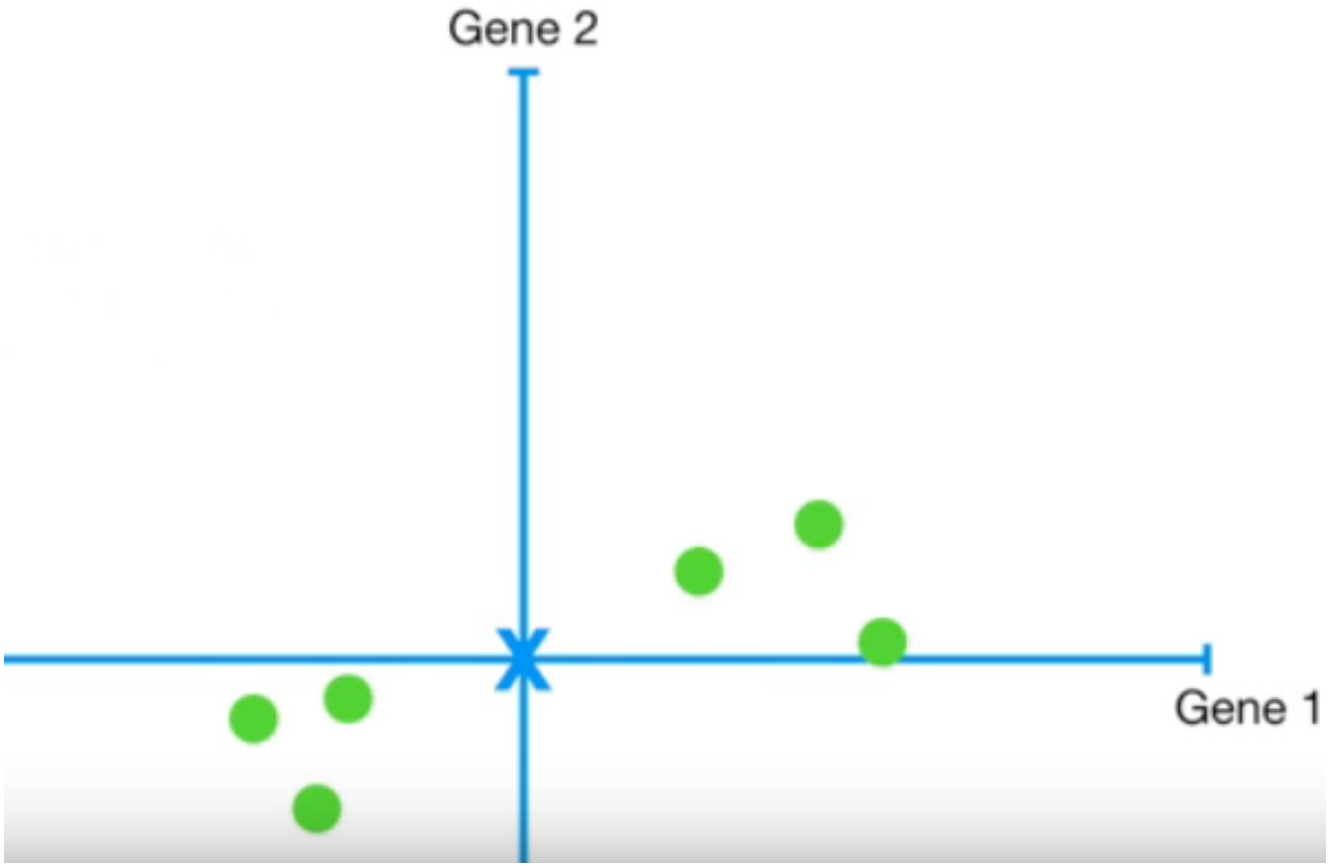
Wyznaczanie średnich dla wierszy



$$u[m] = \frac{1}{N} \sum_{n=1}^{N} X[m, n]$$

PCA - Krok po kroku

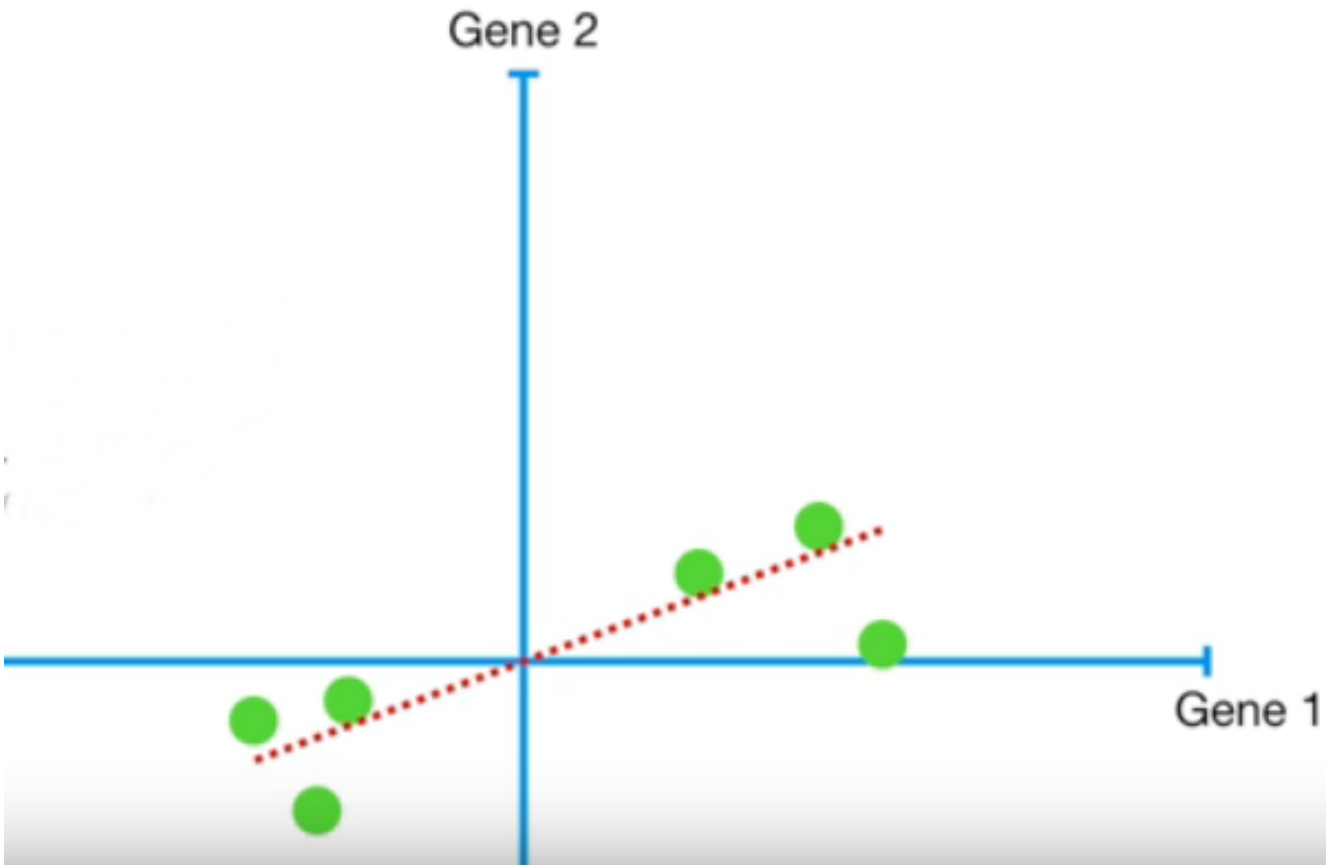
Wyliczanie macierzy odchyleń



$$X'[i,j] = X[i,j] - u[i]$$

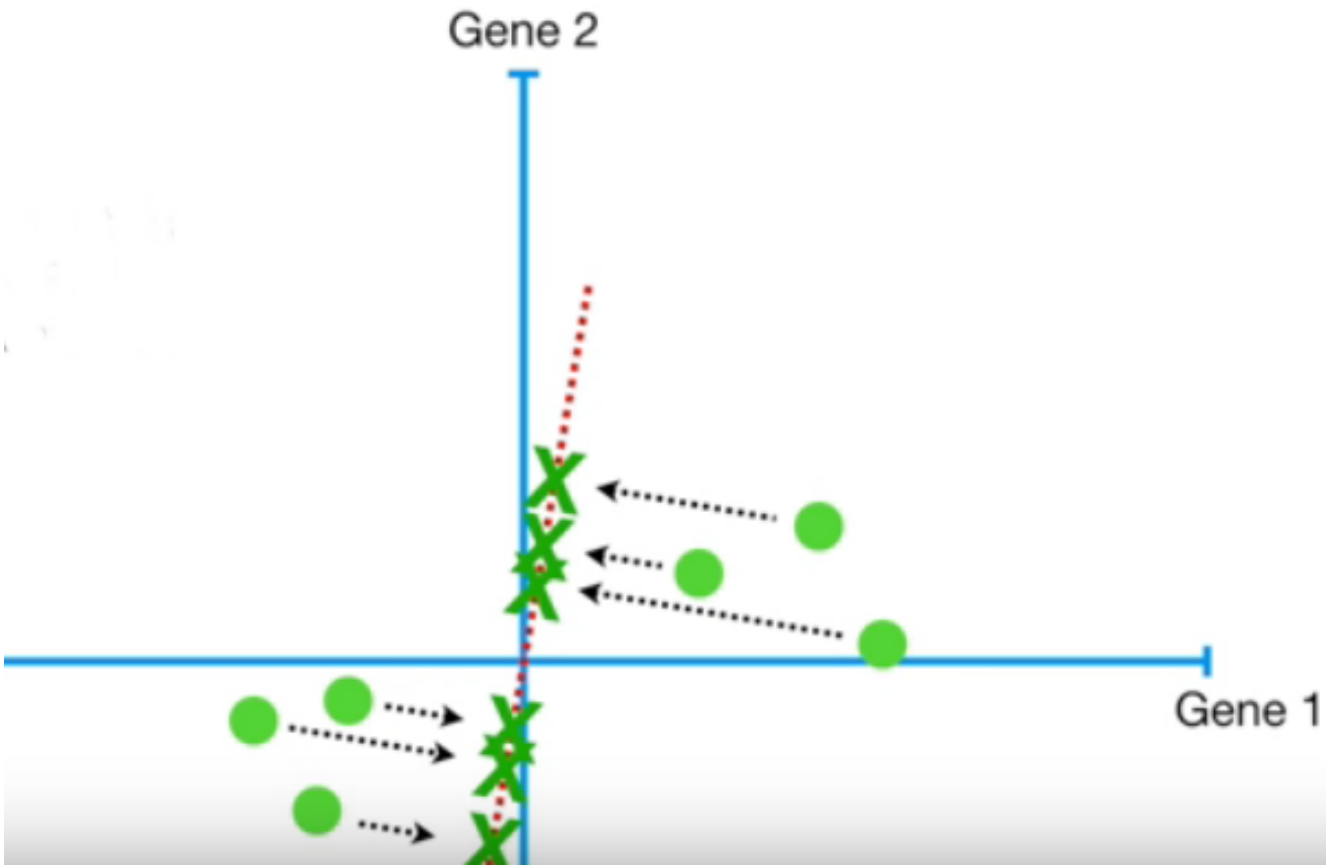
youtube.com/StatQuest

PCA - Krok po kroku



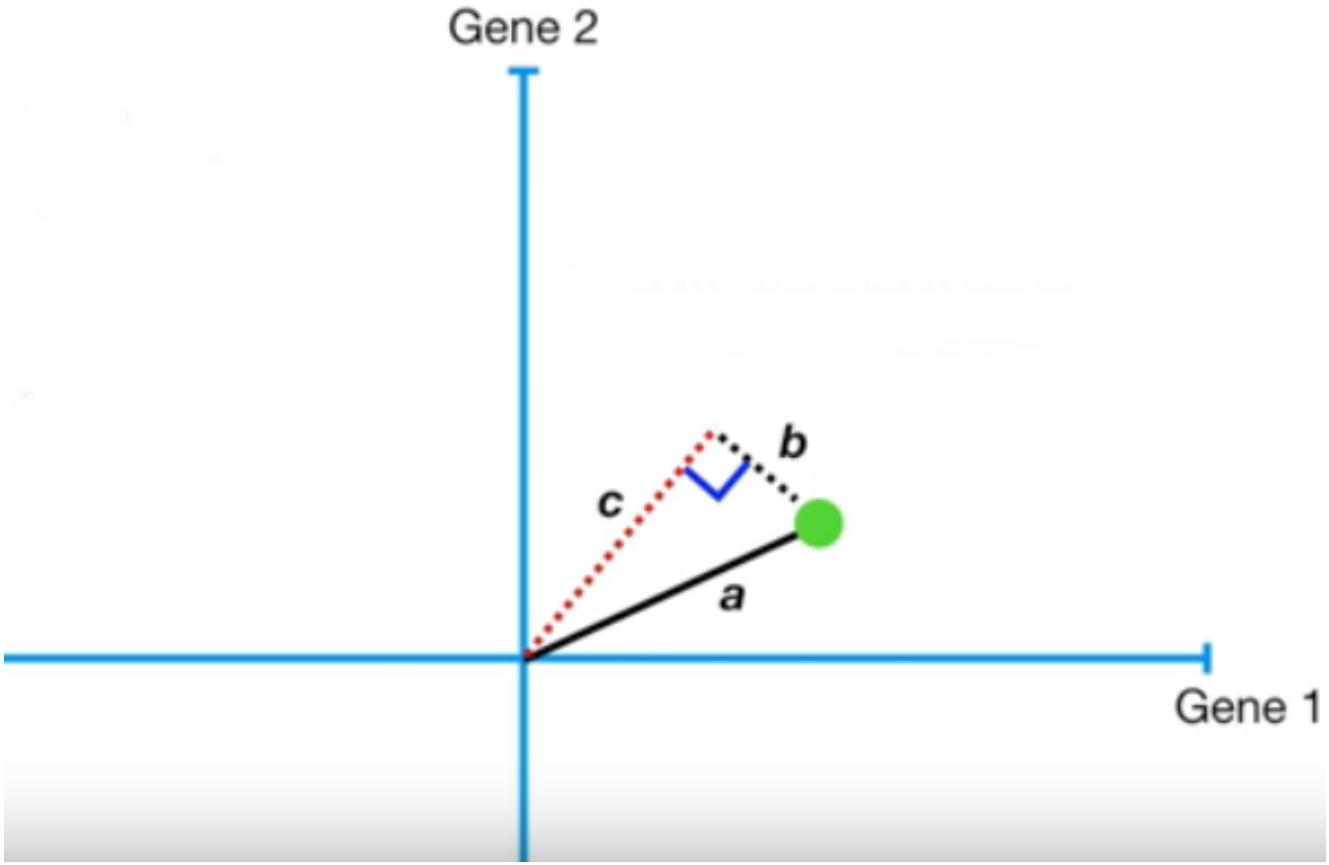
youtube.com/StatQuest

PCA - Krok po kroku



youtube.com/StatQuest

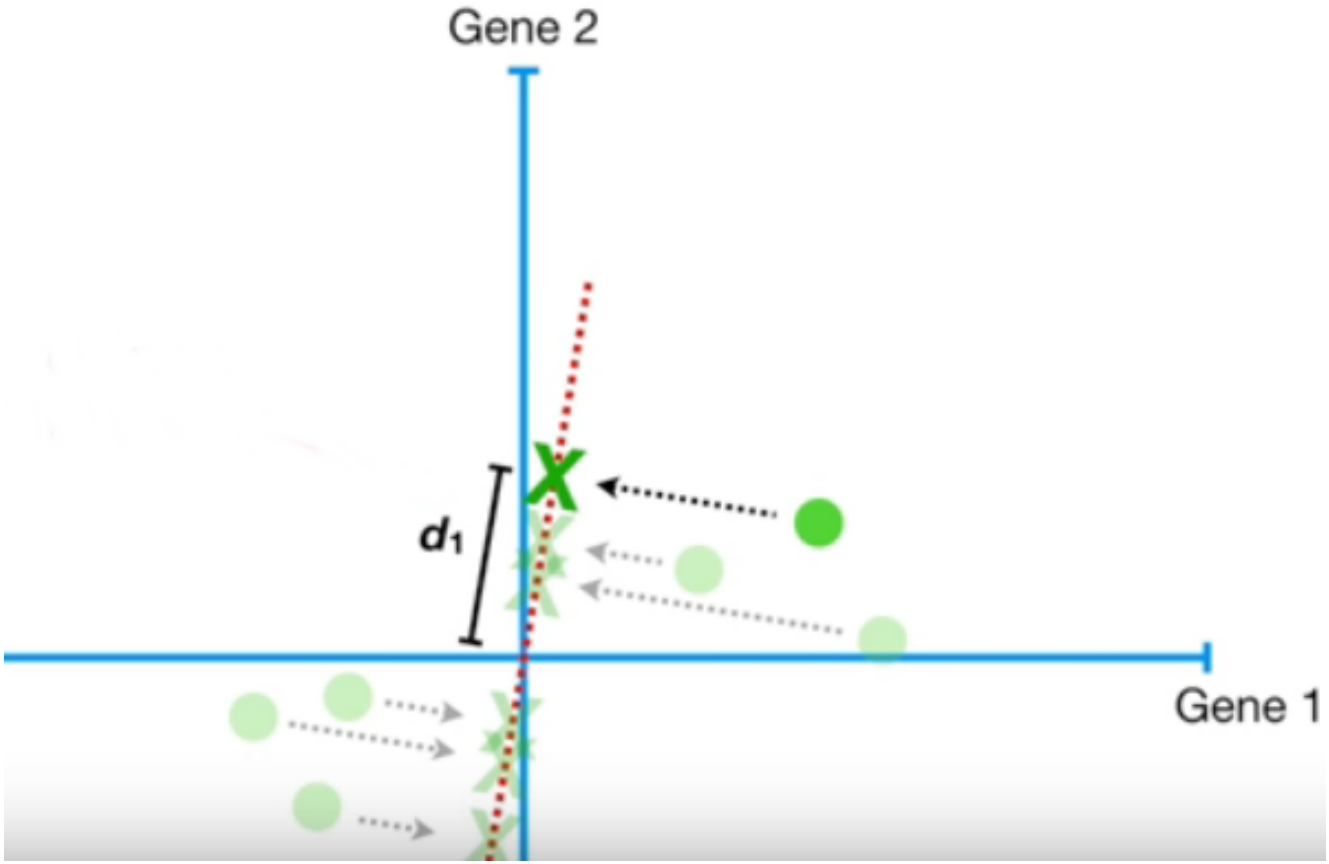
PCA - Krok po kroku



youtube.com/StatQuest

PCA - Krok po kroku

Wartości własne macierzy kowariancji (korelacji)

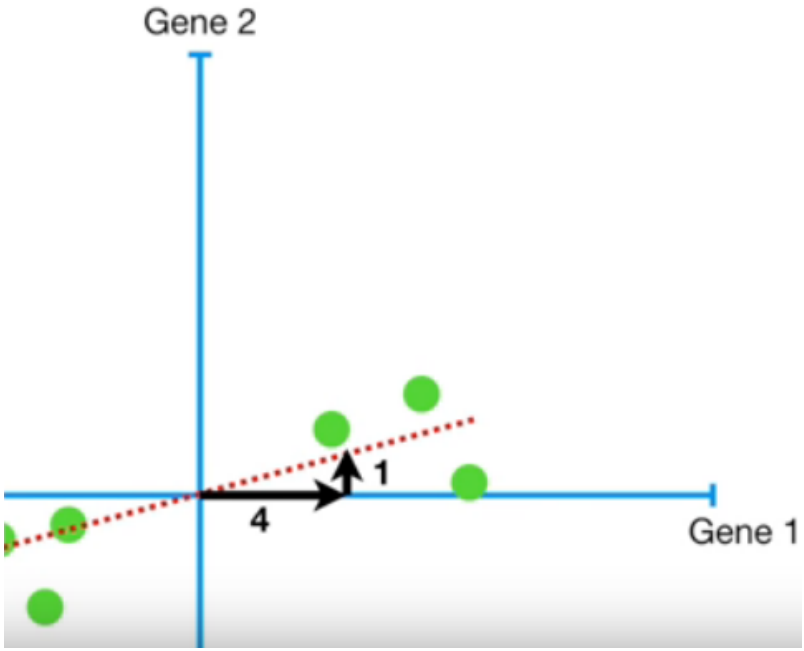


max: $\sum_{i=1}^{N} d_i^2$ - Wartości własne λ

youtube.com/StatQuest

PCA - Krok po kroku

PC1 - 1. składowa główna - Wektory własne ((Eigenvector, Singular Vector) - wektory o jednostkowej długości



$$PC1 = 4x + 1y = 0,97x + 0,24y$$

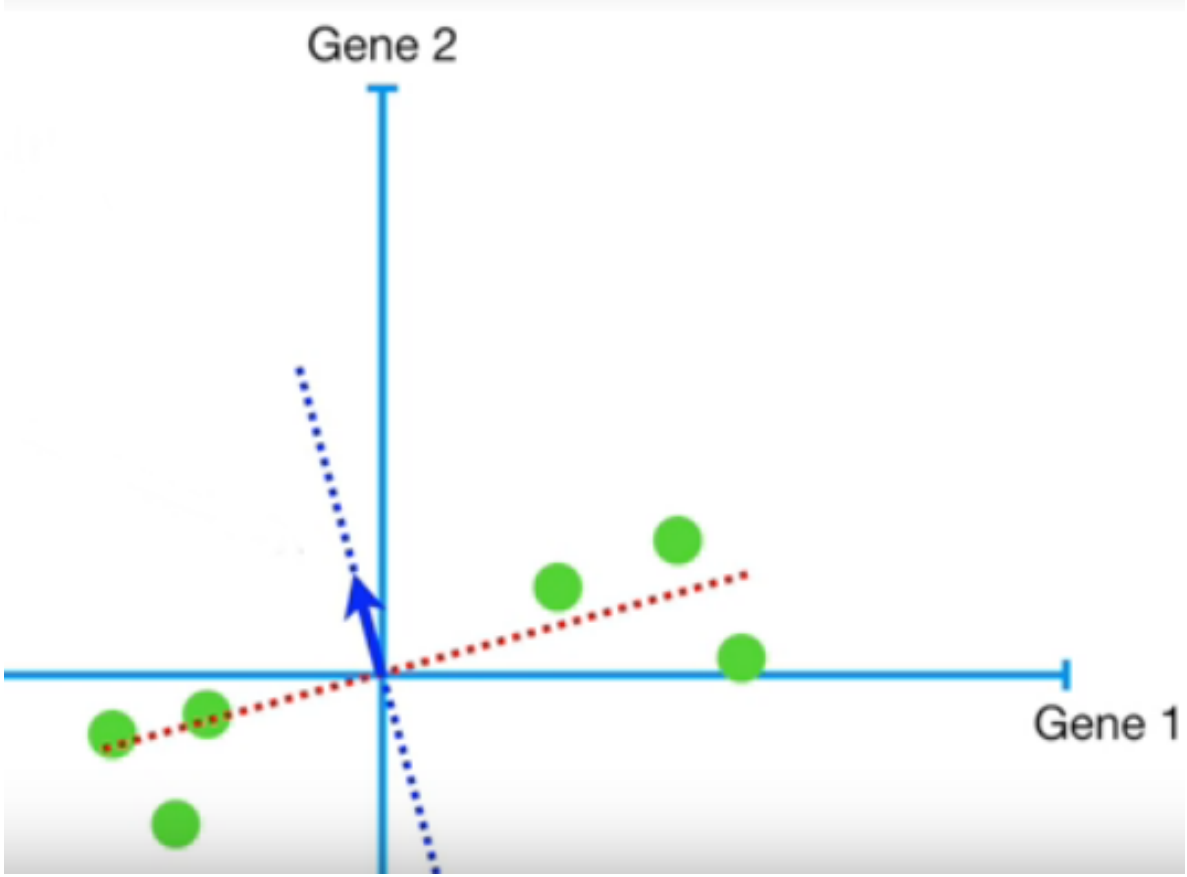
$$PC1 = a_{11}X_1 + a_{12}X2 + ... + a_{1p}X_p$$

a_{ji} - ładunki czynnikowe

youtube.com/StatQuest

PCA - Krok po kroku

PC2



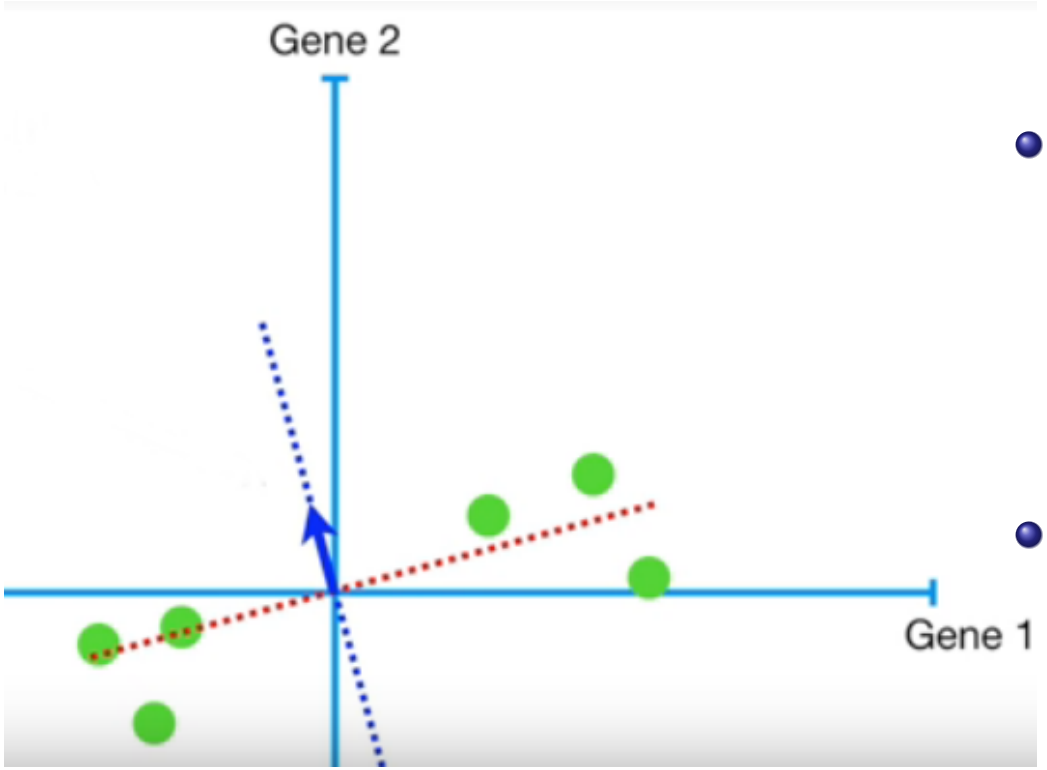
$$PC2 = -x + 4y = -0,24x + 0,97y$$

$$PC2 = a_{21}X_1 + a_{22}X2 + ... + a_{2p}X_p$$

youtube.com/StatQuest

PCA - Krok po kroku

PCi

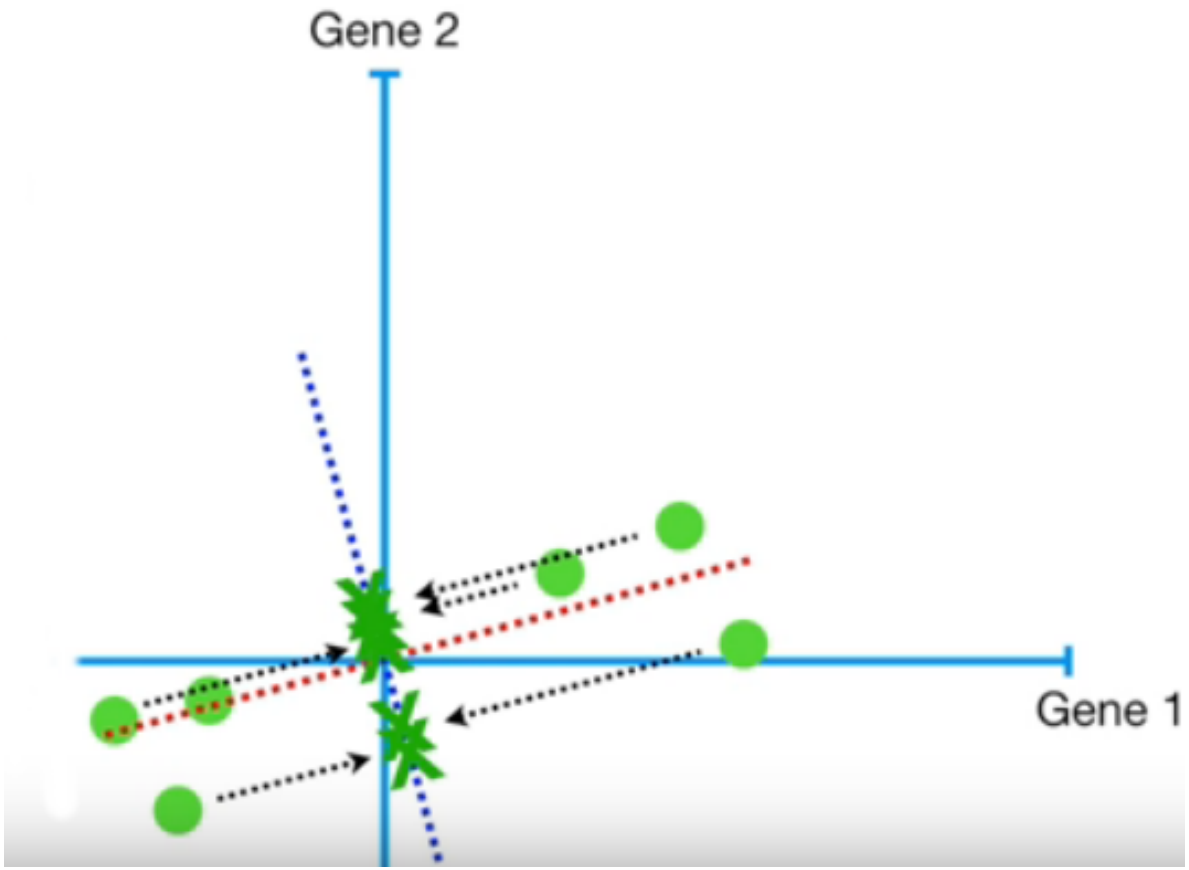


- Kolejna składowa główna jest zdefiniowana tak, aby maksymalizować zmienność, która nie została wyjaśniona przez poprzednie składowe,
- Składowe są wzajemnie ortogonalne, wzajemnie nieskorelowane

$$PCi = a_{i1}X_1 + a_{i2}X2 + ... + a_{ip}X_p$$

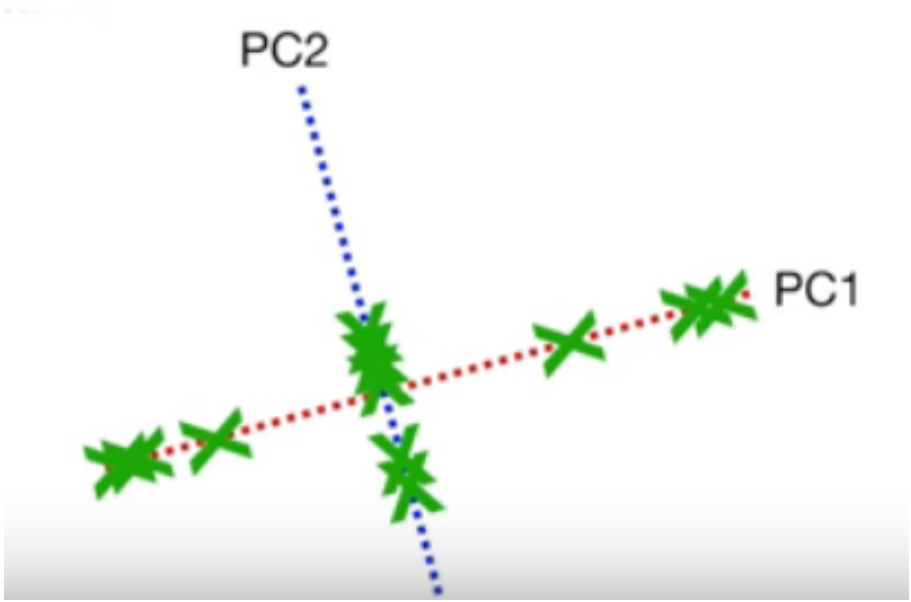
youtube.com/StatQuest

PCA - Krok po kroku



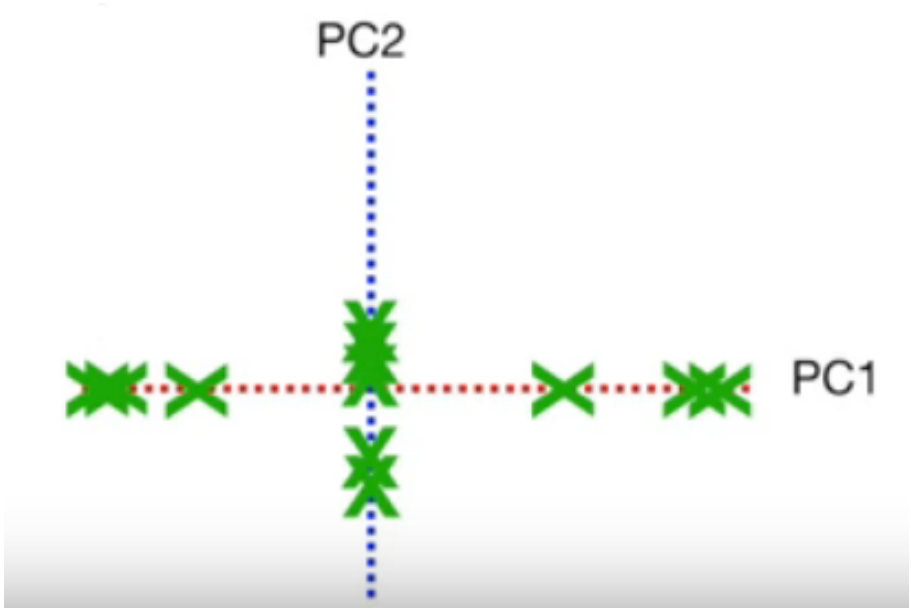
youtube.com/StatQuest

PCA - Krok po kroku



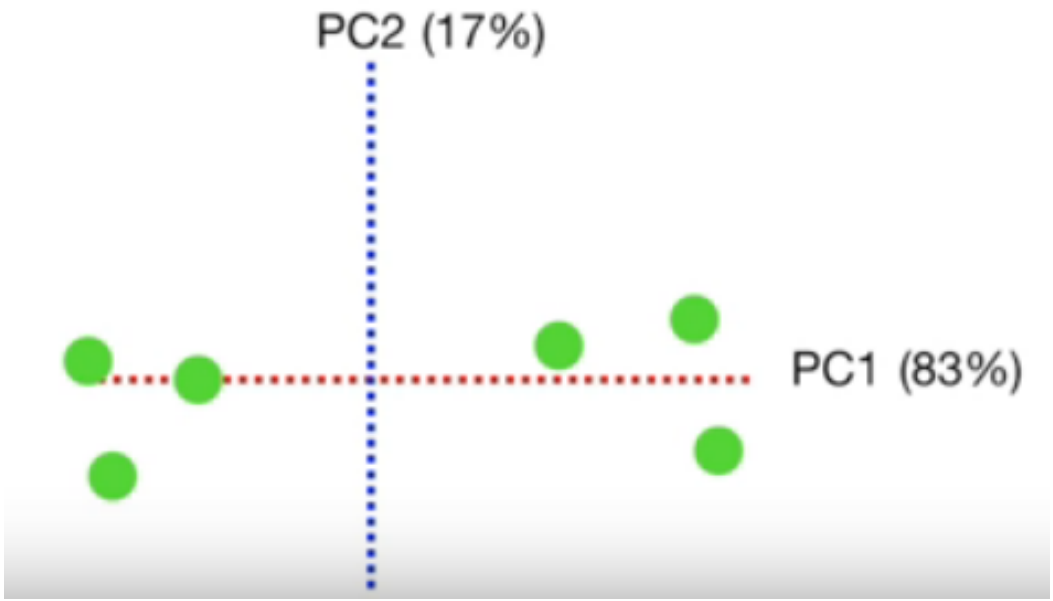
youtube.com/StatQuest

PCA - Krok po kroku



youtube.com/StatQuest

PCA - Krok po kroku



$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + ... + \lambda_p} \cdot 100\%$$

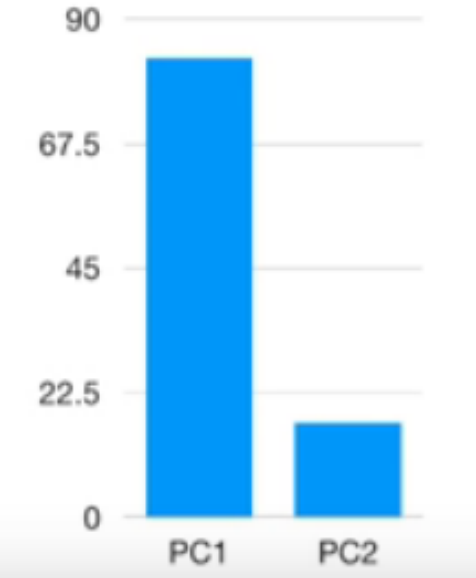
gdzie λ - Eigenvalue,

$\lambda = \sum_{i=1}^{N} d_i^2$

youtube.com/StatQuest

PCA - Krok po kroku

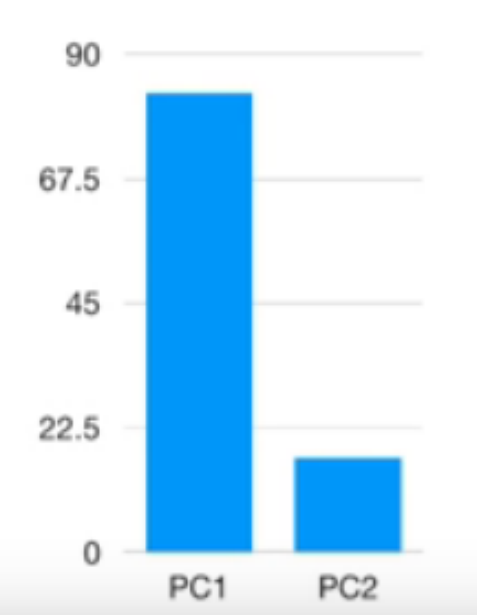
Scree plot

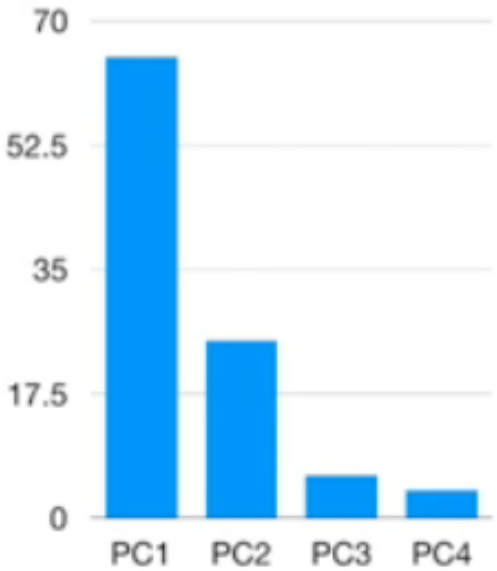


youtube.com/StatQuest

PCA - Krok po kroku

Scree plot

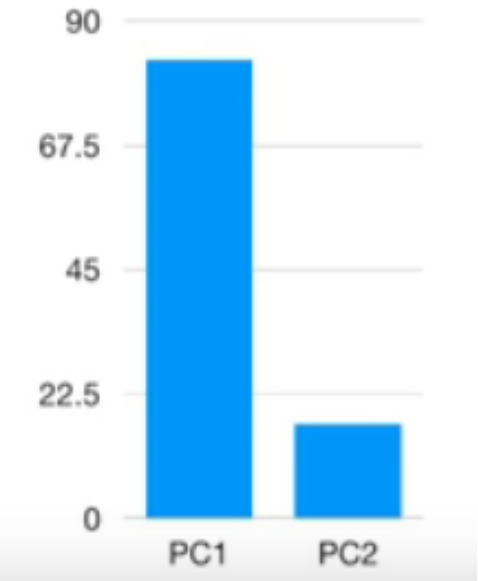


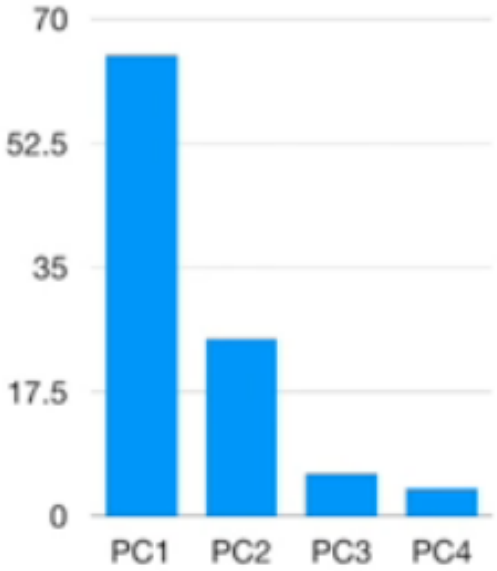


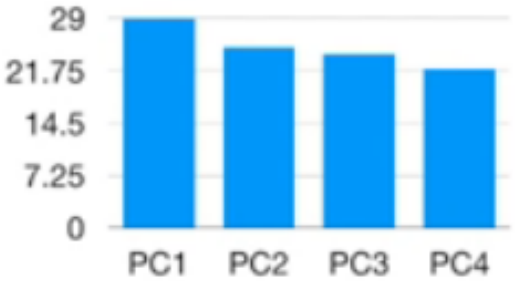
youtube.com/StatQuest

PCA - Krok po kroku

Scree plot

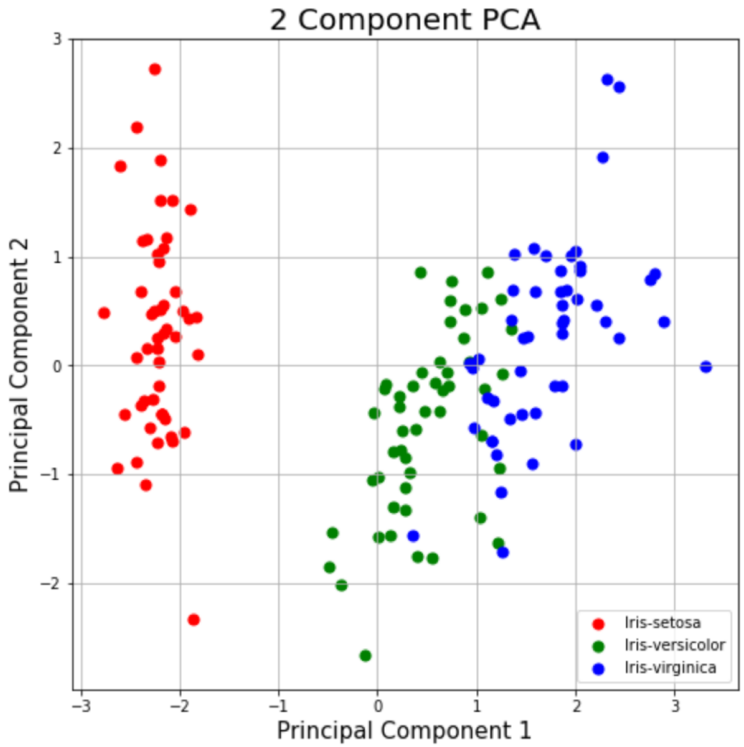


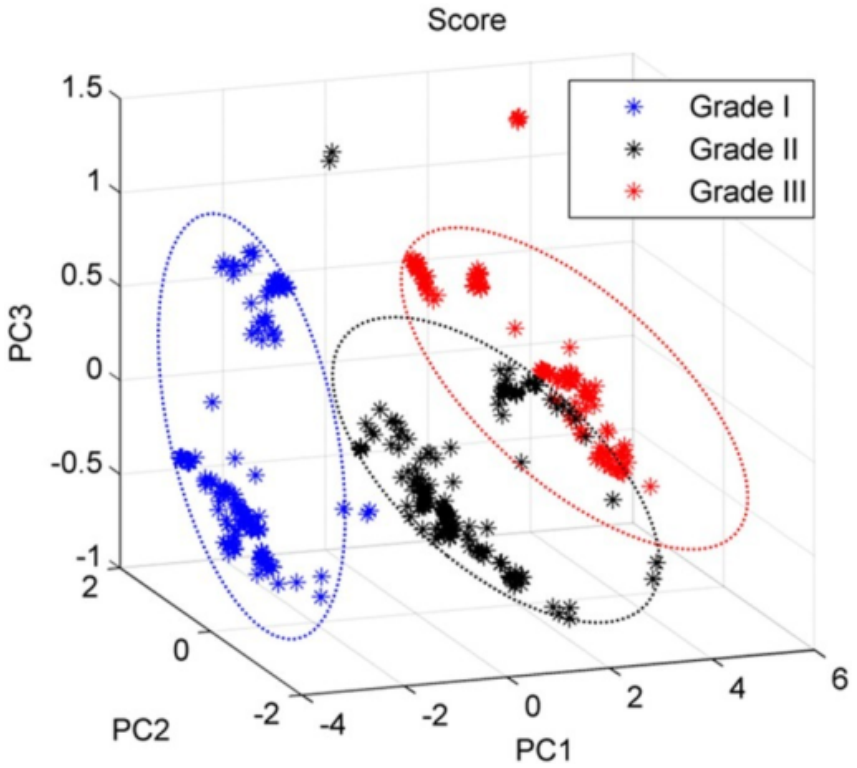




youtube.com/StatQuest

PCA - Przykłady





Kumar R, Singh GP, Gronhaug KM, Afseth NK, de Lange Davies C, Drogset JO, Lilledahl MB, Single cell confocal Raman spectroscopy of human osteoarthritic chondrocytes: a preliminary study, Int J Mol Sci, 2015

ICA - Independent Component Analysis (Metoda składowych niezależnych)

Liniowy algorytm separujący dane wejściowe na komponenty (nie-gaussowskie) statystycznie niezależne.

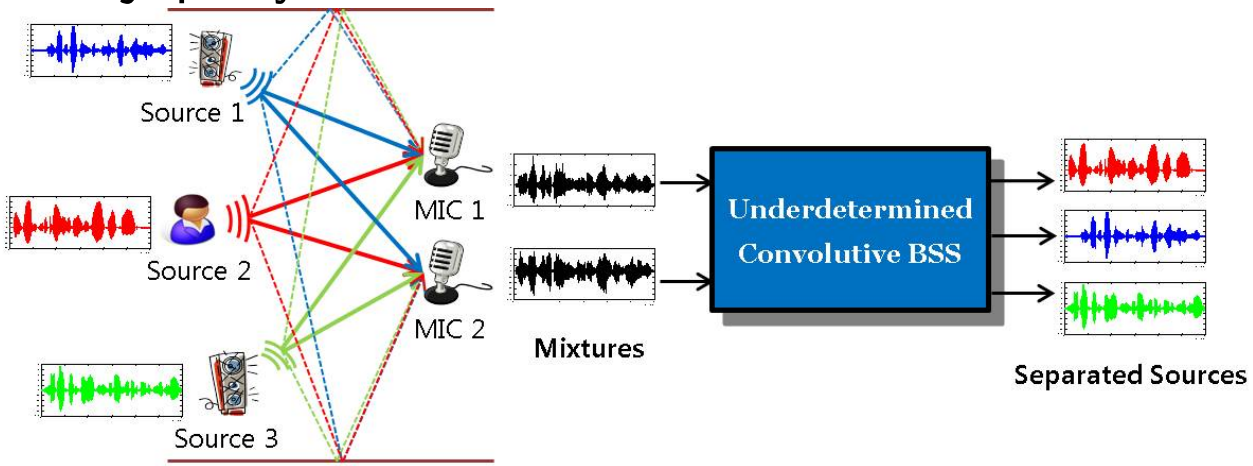
$$P(Y_1, Y_2) = P(Y_1) \cdot P(Y_2)$$

ICA - Independent Component Analysis (Metoda składowych niezależnych)

Liniowy algorytm separujący dane wejściowe na komponenty (nie-gaussowskie) statystycznie niezależne.

$$P(Y_1, Y_2) = P(Y_1) \cdot P(Y_2)$$

Jest to przykład Blind Source Separation, rozwiązanie problemu koktajl party.



SangGyun Kim, Chang D. Yoo, Machine Learning for Speech Processing,Blind Source Separation (BSS)

ICA - Definicja

Dane reprezentowane przez wektor obserwacji

$$\mathbf{x} = (x_1, ..., x_m)^T$$

Wektor ukrytych komponentów

$$\mathbf{s} = (s_1, ..., s_n)^T.$$

ICA - Definicja

Dane reprezentowane przez wektor obserwacji

$$\mathbf{x} = (x_1, ..., x_m)^T$$

Wektor ukrytych komponentów

$$\mathbf{s} = (s_1, ..., s_n)^T.$$

$$x_i = a_{i,1}s_1 + ... + a_{i,n}s_n$$

$$\mathbf{x} = \mathbf{As}$$

ICA - Definicja

Dane reprezentowane przez wektor obserwacji

$$\mathbf{x} = (x_1, ..., x_m)^T$$

Wektor ukrytych komponentów

$$\mathbf{s} = (s_1, ..., s_n)^T.$$

$$x_i = a_{i,1}s_1 + ... + a_{i,n}s_n$$

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

Zadaniem jest transformacja wektora $\mathbf{x}$ w wektor $\mathbf{s}$ (właściwie $\mathbf{y}$) przy pomocy macierzy separującej $\mathbf{W} = \mathbf{A}^{-1}$, gdzie komponenty wektora $\mathbf{s}$ są maksymalnie niezależne (mierzone funkcją niezależności $F(s_1, ..., s_n)$).

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

ICA - Algorytm optymalizacji

Problem optymalizacji - maksymalizacja statystycznej niezależności komponentów:

- Minimalizacja **wzajemnej informacji** (mutual information) wynikowych komponentów,
  Informacja wzajemna - miara tego, o ile poznanie jednej zmiennej zmniejsza niepewność o drugiej. Zerowa informacja wzajemna to zmienne niezależne.
  Używane miary:
  - Dywergencja Kullbacka-Leiblera,
  - maksymalna entropia.
- Maksymalizacja nie-gaussowości - na podstawie centralnego twierdzenia granicznego.
  Używane miary:
  - kurtoza - miara spłaszczenia rozkładu,
  - negentropia - negatywna entropia.

ICA - założenia

Osiągnięcie dobrego wyniku ICA wymaga założeń:

1. dane źródłowe są od siebie niezależne,
2. wartości danych źródłowych mają rozkład nie-gaussowski,
3. liczba obserwacji nie może być mniejsza niż liczba źródeł $m \geqslant n$

ICA - algorytm

Typowy algorytm:

1. centrowanie,

ICA - algorytm

Typowy algorytm:

1. centrowanie,
2. whitening - transformacja wektora zmiennych o znanej macierzy kowariancji w nowy wektor, którego macierz kowariancji to macierz jednostkowa (zmienne nieskorelowane o wariancji=1),

ICA - algorytm

Typowy algorytm:

1. centrowanie,
2. whitening - transformacja wektora zmiennych o znanej macierzy kowariancji w nowy wektor, którego macierz kowariancji to macierz jednostkowa (zmienne nieskorelowane o wariancji=1),
3. redukcja wymiaru (PCA),

ICA - algorytm

Typowy algorytm:

1. centrowanie,
2. whitening - transformacja wektora zmiennych o znanej macierzy kowariancji w nowy wektor, którego macierz kowariancji to macierz jednostkowa (zmienne nieskorelowane o wariancji=1),
3. redukcja wymiaru (PCA),
4. maksymalizacja statystycznej niezależności.

PCA vs. ICA

- PC vs. ICA - różne założenia, różne wyniki, jeden cel: wyodrębnienie "oryginalnych" danych

PCA vs. ICA

- PC vs. ICA - różne założenia, różne wyniki, jeden cel: wyodrębnienie "oryginalnych" danych
- PCA nie sprawdza się w BSS,

PCA vs. ICA

- PC vs. ICA - różne założenia, różne wyniki, jeden cel: wyodrębnienie "oryginalnych" danych
- PCA nie sprawdza się w BSS,
- Wyniki ICA są zależne od struktury macierzy danych,

PCA vs. ICA

- PC vs. ICA - różne założenia, różne wyniki, jeden cel: wyodrębnienie "oryginalnych" danych
- PCA nie sprawdza się w BSS,
- Wyniki ICA są zależne od struktury macierzy danych,
- ICA nie jest w stanie odtworzyć ilości sygnałów źródłowych,
- PCA skupia się na globalnych cechach, ICA na elementach składowych danych,

PCA vs. ICA

- PC vs. ICA - różne założenia, różne wyniki, jeden cel: wyodrębnienie "oryginalnych" danych
- PCA nie sprawdza się w BSS,
- Wyniki ICA są zależne od struktury macierzy danych,
- ICA nie jest w stanie odtworzyć ilości sygnałów źródłowych,
- PCA skupia się na globalnych cechach, ICA na elementach składowych danych,
- wyniki ICA są wzajemnie niezależne, PCA wzajemnie ortogonalne.

Funkcje kosztu i optymalizacja

Funkcja kosztu - funkcja określająca w jaki sposób "karzemy" algorytm za błędy; sposób oceny, na ile algorytm się myli.
Optymalizacja - procedura efektywnego wyszukiwania najlepszego rozwiązania, często poprzez minimalizację funkcji kosztu (błędu algorytmu).

Funkcje kosztu

- Funkcje kosztu dla regresji:

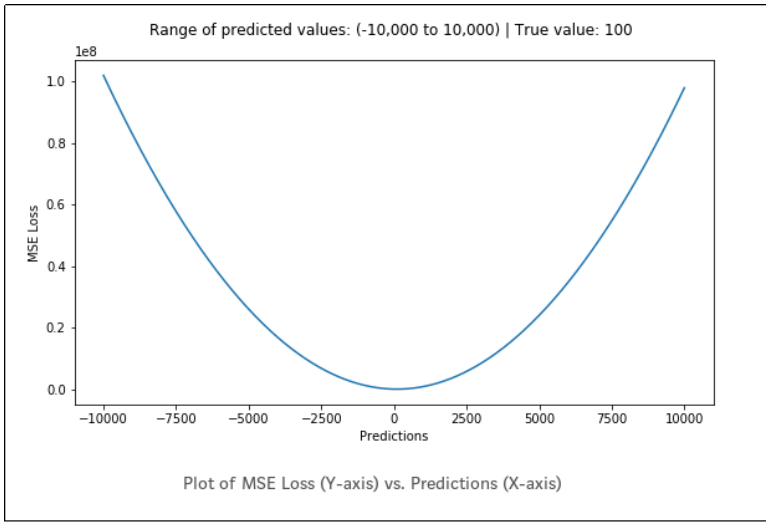
- Funkcje kosztu dla klasyfikacji (wynikiem jest prawdopodobieństwo)

Funkcje kosztu

- Funkcje kosztu dla regresji:
  - błąd średniokwadratowy (Mean Square Error)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2$$

gdzie:
N - liczba danych,
f_i - wartość zwrócona przez model,
y_i - wartość rzeczywista

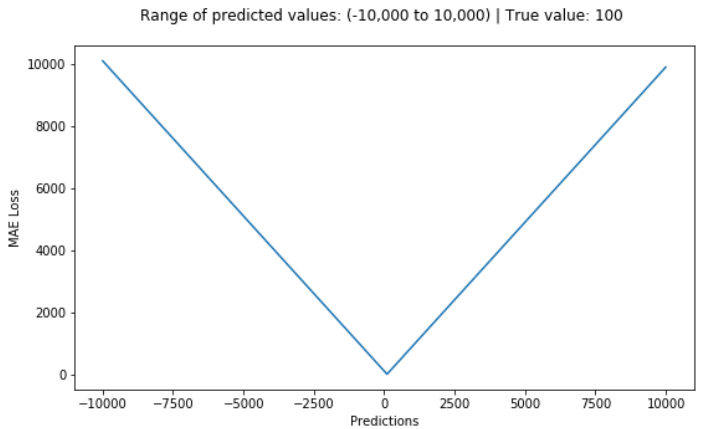


Funkcje kosztu

- Funkcje kosztu dla regresji:
  - błąd średniokwadratowy (Mean Square Error)
  - Mean Absolute Error

$$MAE = \frac{1}{N}\sum_{i=1}^{n}|f_i - y_i|2$$

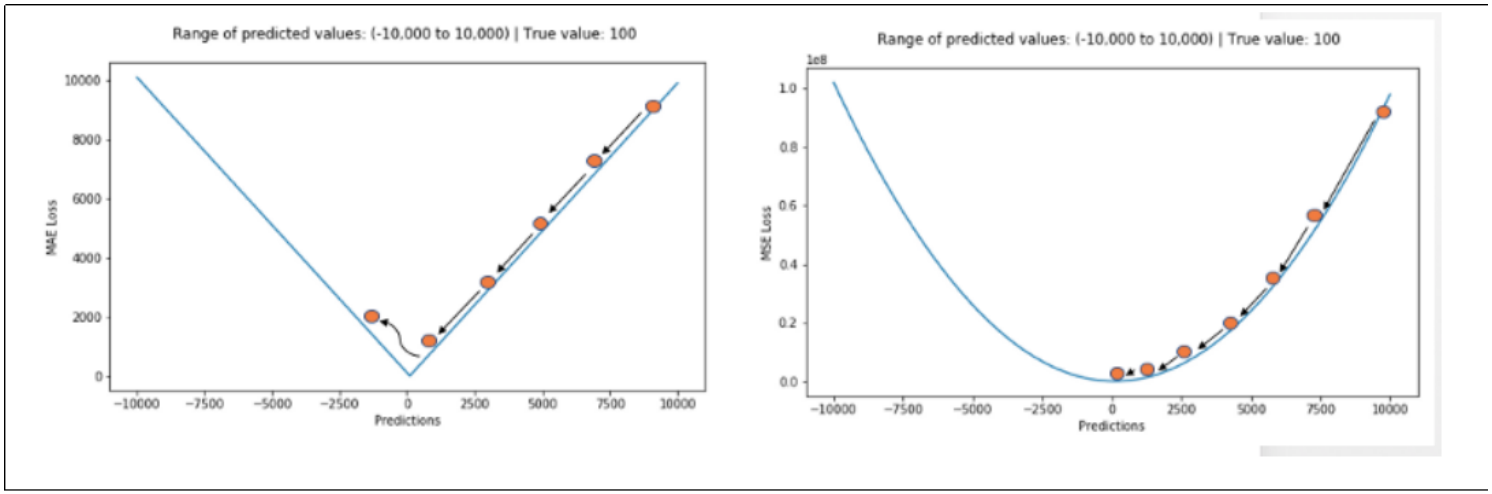
gdzie:
N - liczba danych,
f_i - wartość zwrócona przez model,
y_i - wartość rzeczywista



Plot of MAE Loss (Y-axis) vs. Predictions (X-axis)

Funkcje kosztu

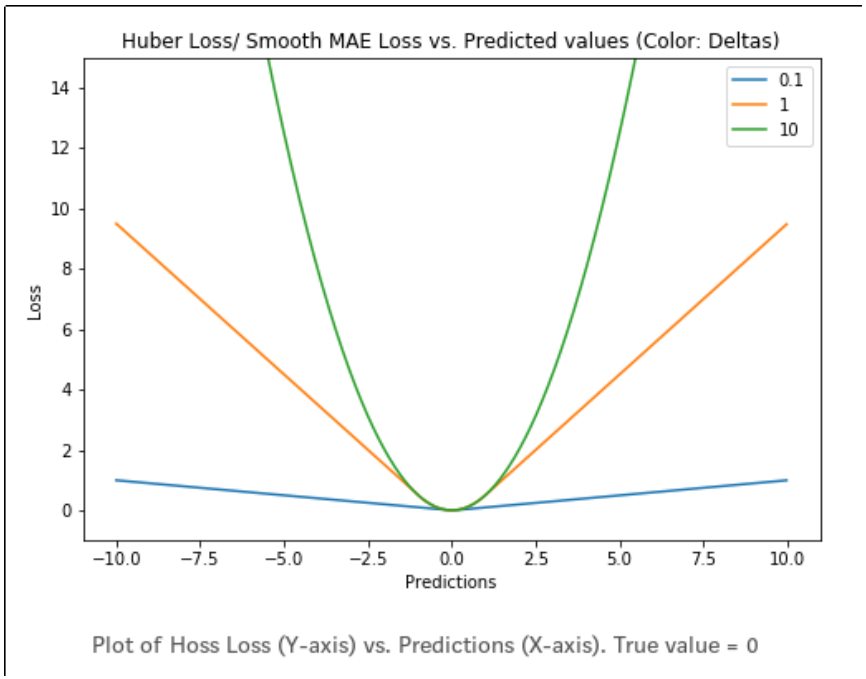
- Funkcje kosztu dla regresji:
  - błąd średniokwadratowy (Mean Square Error)
  - Mean Absolute Error



Funkcje kosztu

- Funkcje kosztu dla regresji:
  - błąd średniokwadratowy (Mean Square Error)
  - Mean Absolute Error
  - Huber loss

    Kształt zależny od dobieranego parametru δ
    Dla $\delta \backsim 0- > MAE$
    dla $\delta \backsim \infty- > MSE$

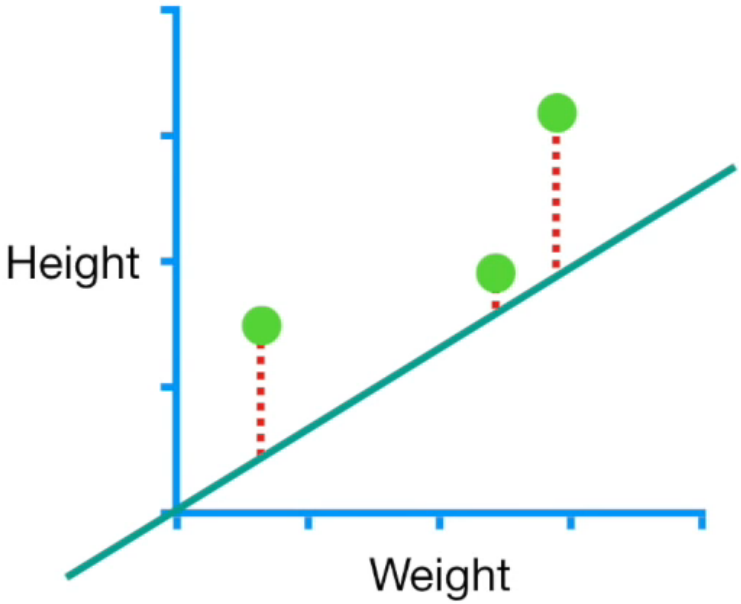


Plot of Hoss Loss (Y-axis) vs. Predictions (X-axis). True value = 0

Funkcje kosztu

- Funkcje kosztu dla regresji:
  - błąd średniokwadratowy (Mean Square Error)
  - Mean Absolute Error
  - Huber loss
- Funkcje kosztu dla klasyfikacji (wynikiem jest prawdopodobieństwo)
  - Entropia,
  - Entropia krzyżowa (cross entropy),
  - SVM

Optymalizacja

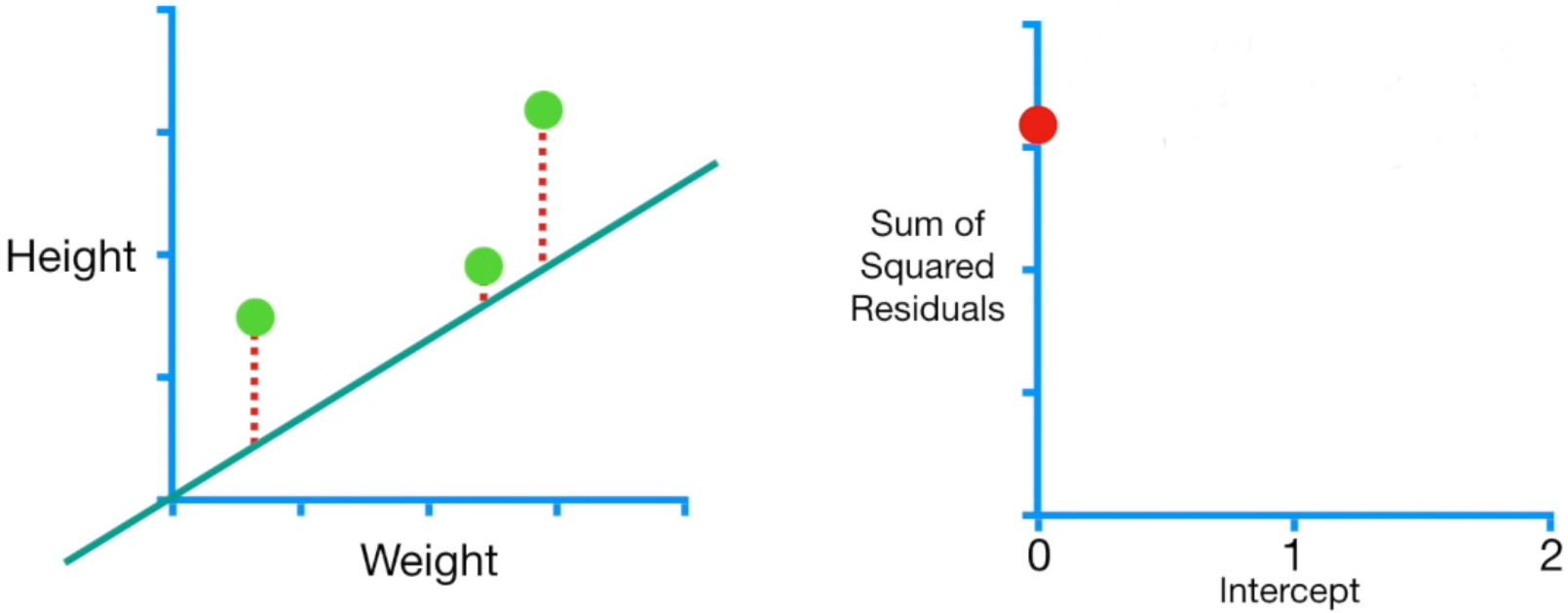


Przykładem optymalizacji jest regresja liniowa - dopasowanie prostej do pomiarów

W tym przypadku szukamy takich parametrów prostej, aby fukncja kosztu była jak najmniejsza

StatQuest

Optymalizacja

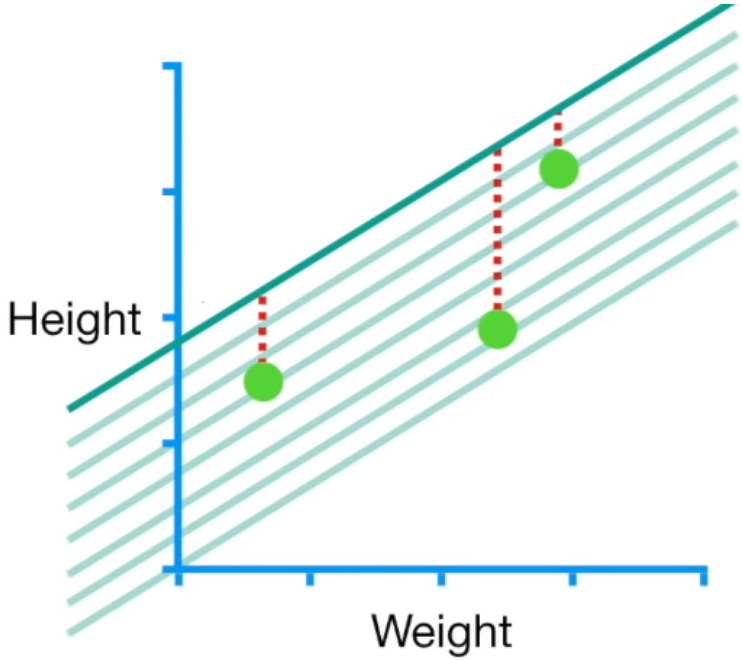


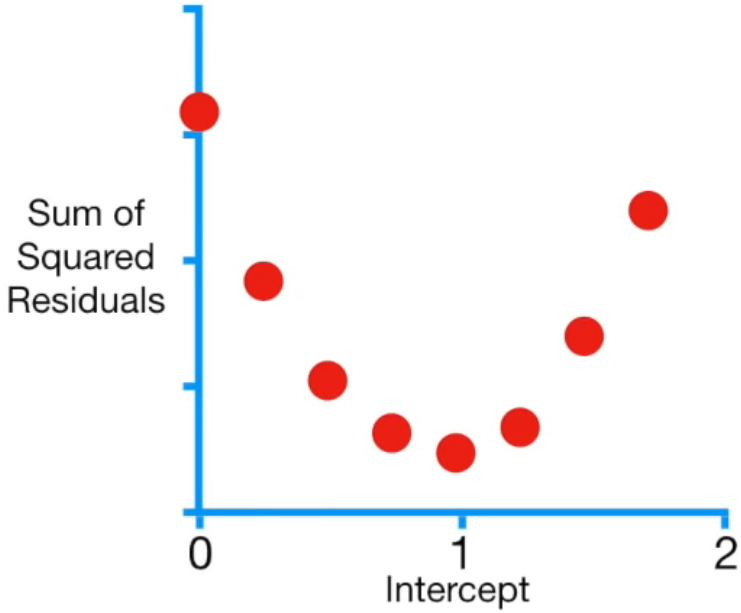
Suma kwadratów różnic pomiarów (jasno-zielone punkty) oraz proponowanej prostej (zielona linia) zaznaczona jest czerwonym punktem

Poszukujemy jak najmniejszej wartości

StatQuest

Optymalizacja



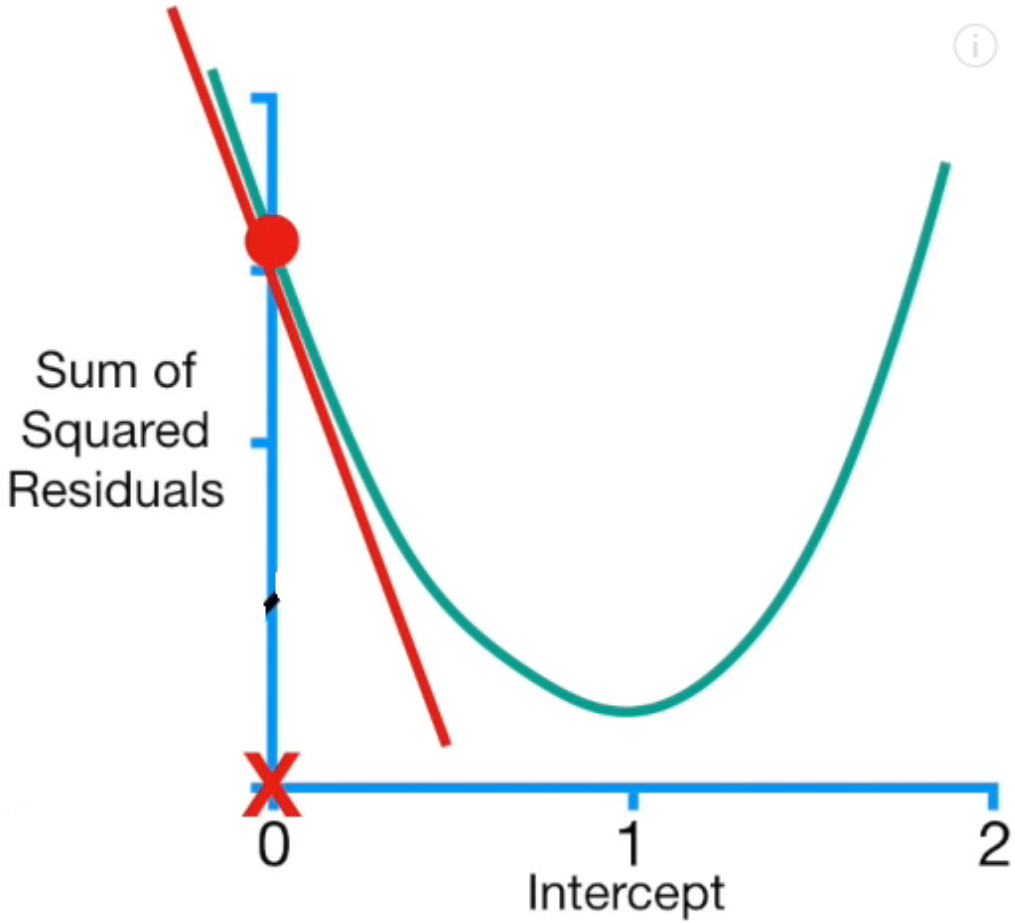


Po kilku propozycjach prostych

Nie mamy pewności, czy punkt o najmniejszej wartości jest w rzeczywistości takim. Aby to sprawdzić musielibyśmy zaproponować i obliczyć błędy dla jeszcze wielu innych prostych

StatQuest

Optymalizacja - Gradient descent

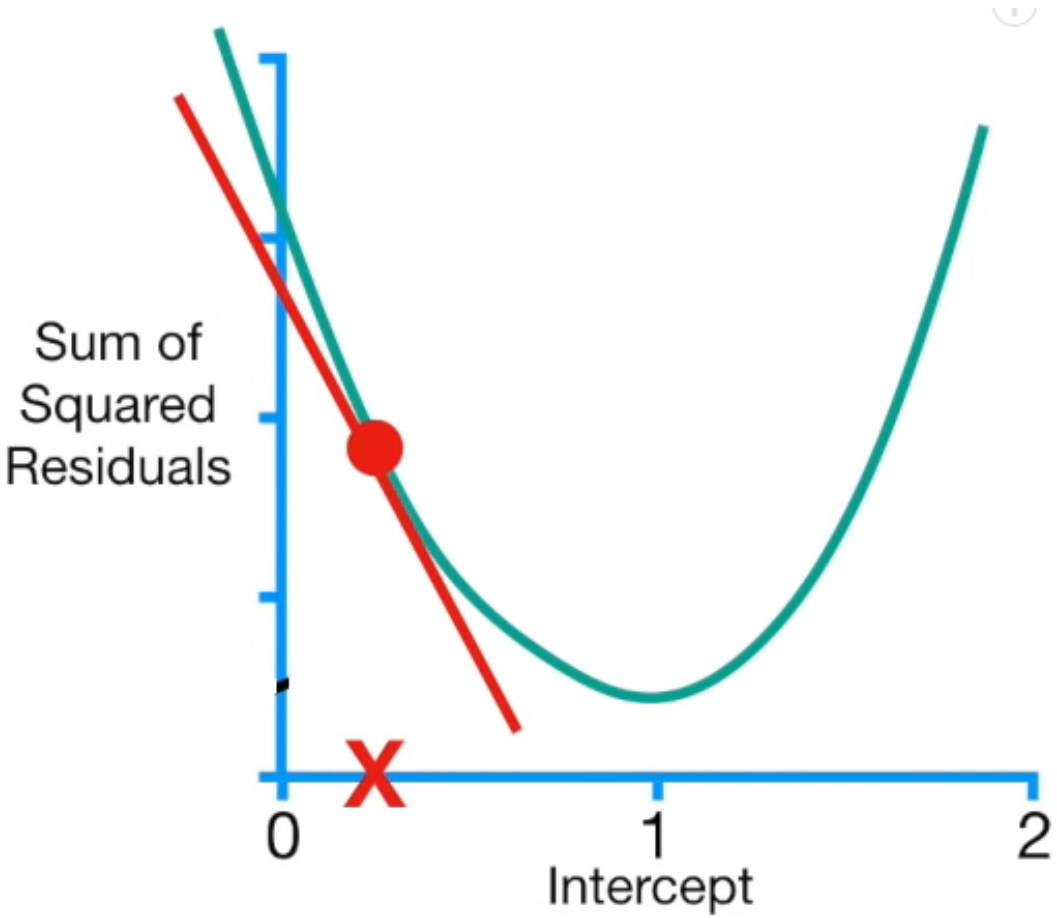


Gradient descent to algorytm, który nie wymaga obliczania funkcji kosztu dla dużej liczby propozycji

Propozycje parametrów (oznaczone czerwonym krzyżykiem) wyznaczane są na podstawie nachylenia zbocza pochodnej funkcji:

StatQuest

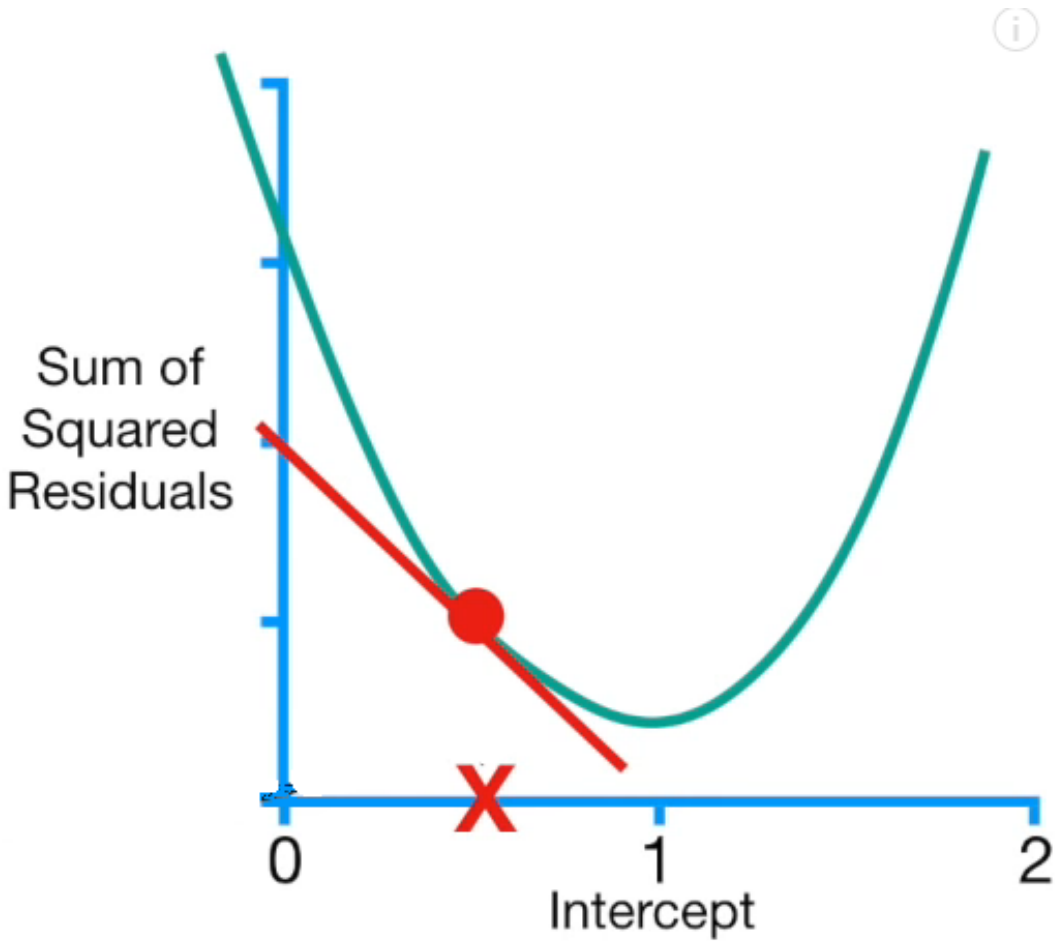
Optymalizacja - Gradient descent



Im większe nachylenie zbocza, tym bardziej oddalona jest następna propozycja

StatQuest

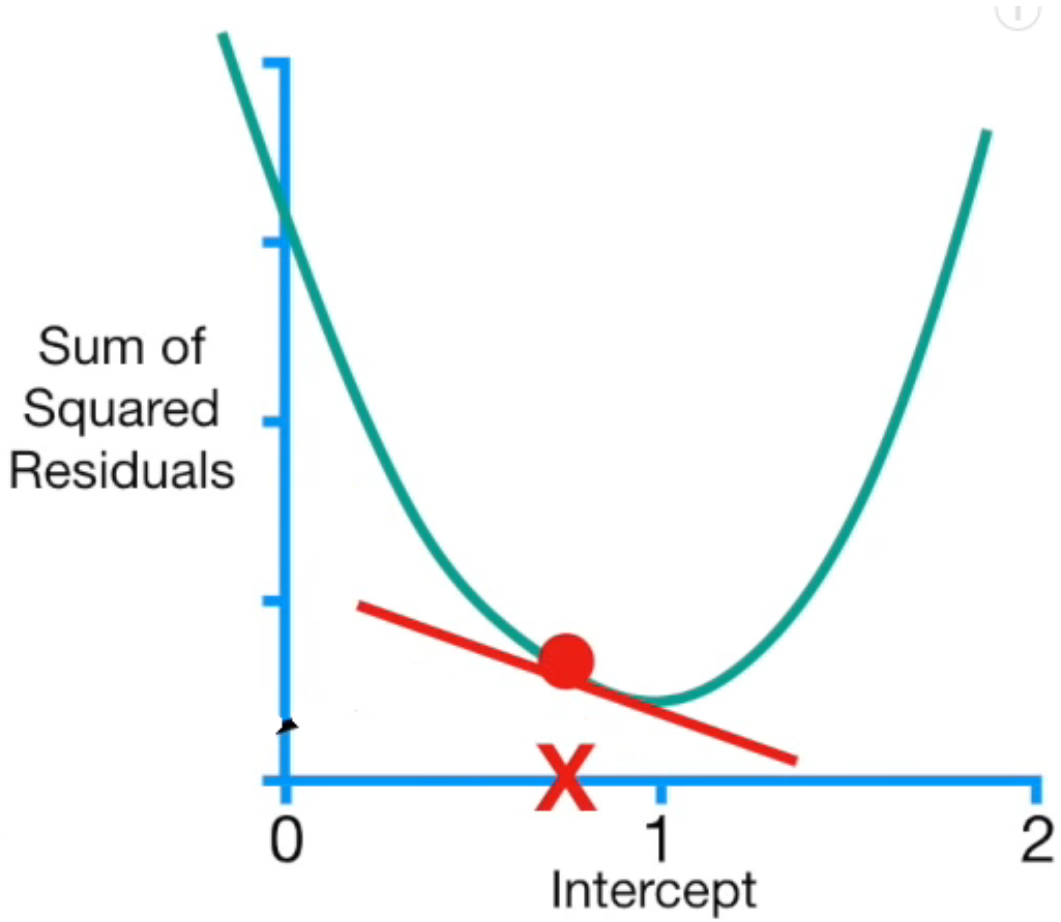
Optymalizacja - Gradient descent



Im mniejsza pochodna, tym mniejsze odstępy między propozycjami

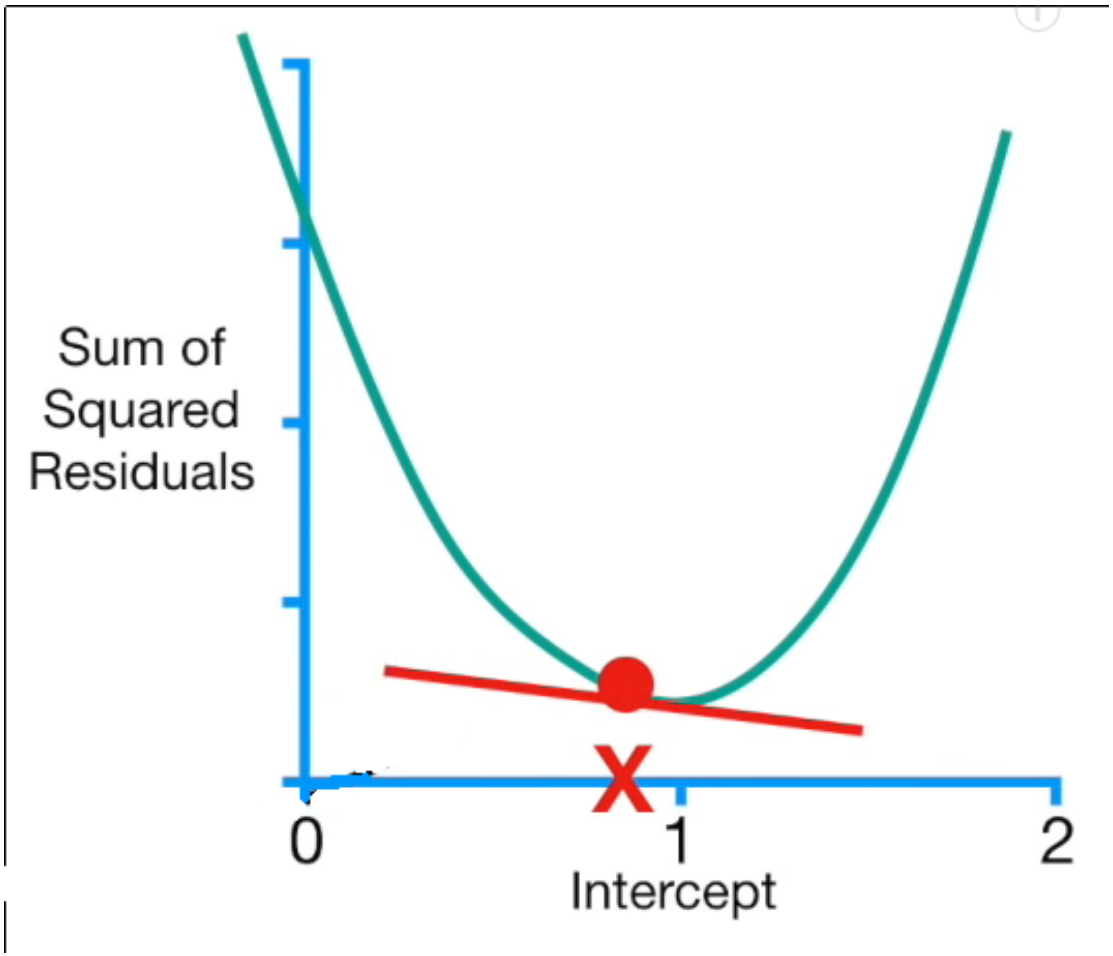
StatQuest

Optymalizacja - Gradient descent



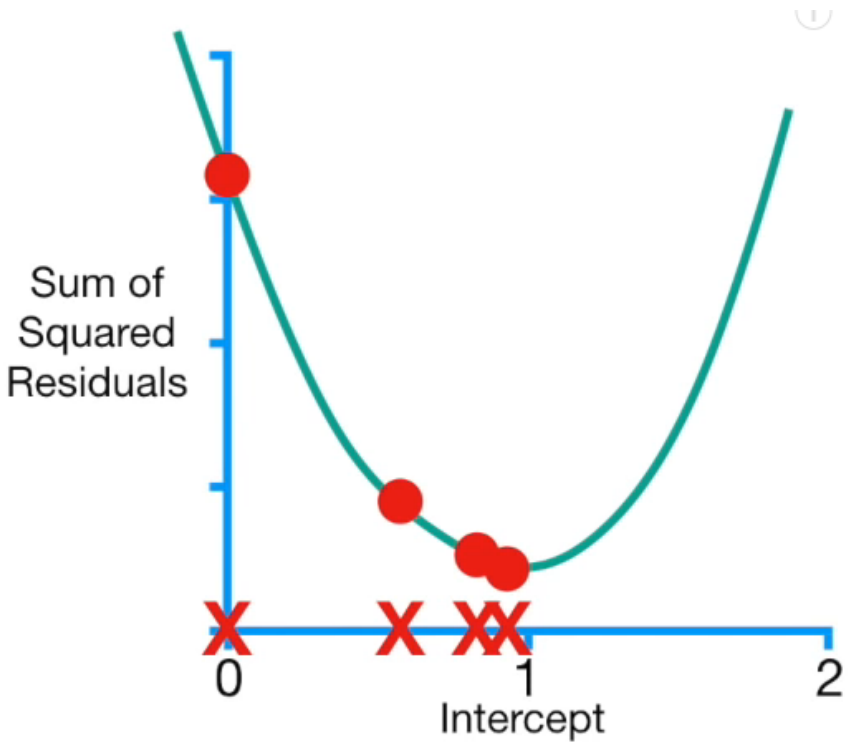
StatQuest

Optymalizacja - Gradient descent



StatQuest

Optymalizacja - Gradient descent



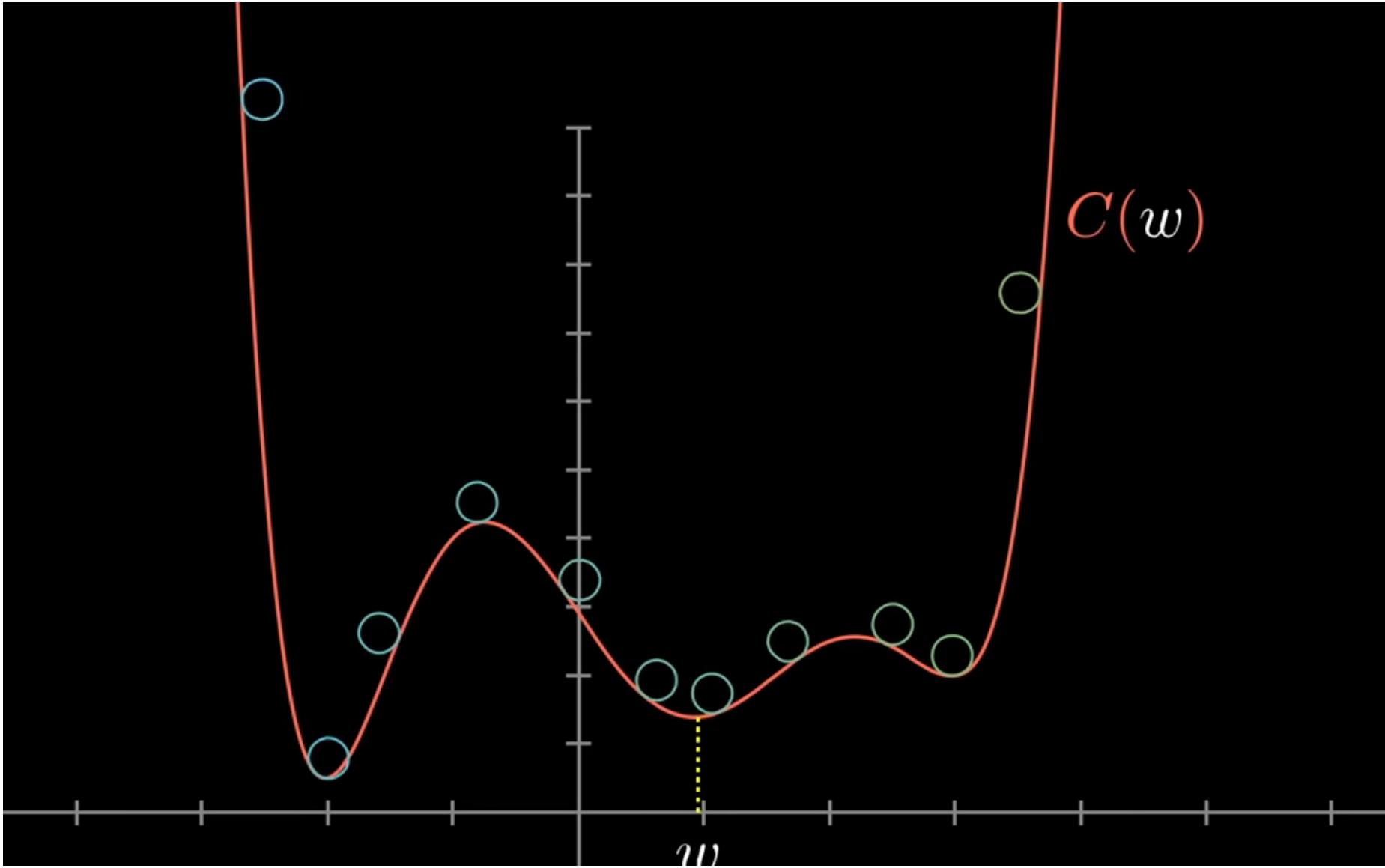
Wielkość kroku zależy od nachylenia zbocza oraz parametru Learning Rate:

Step Size=Slope x Learinng Rate

Warunki zakończenia optymalizacji:

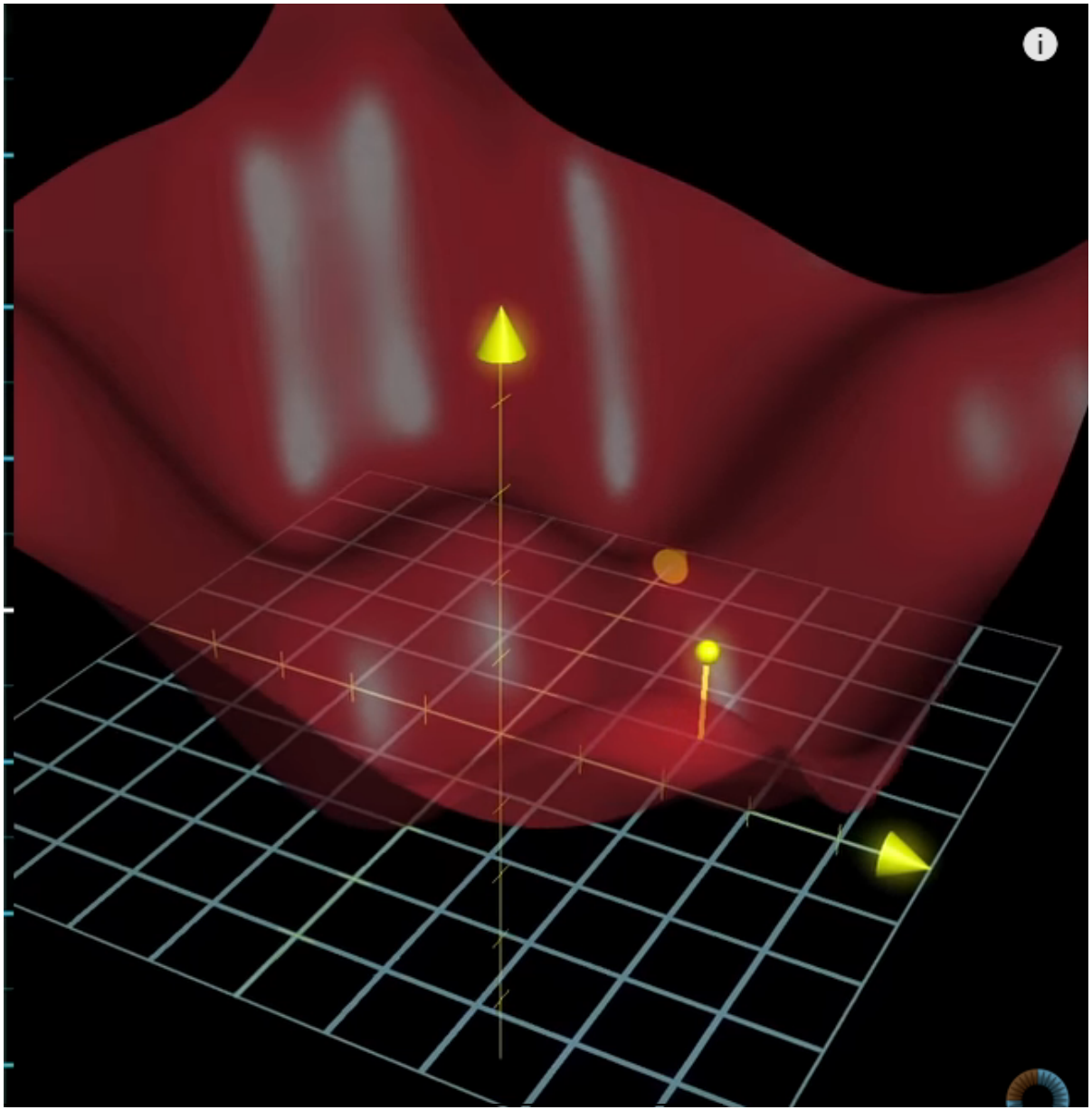
1. wielkośc kroku mniejsza niż założona wartość,
2. ilość kroków (iteracji) większa niż założona wartość.

Optymalizacja - Gradient descent



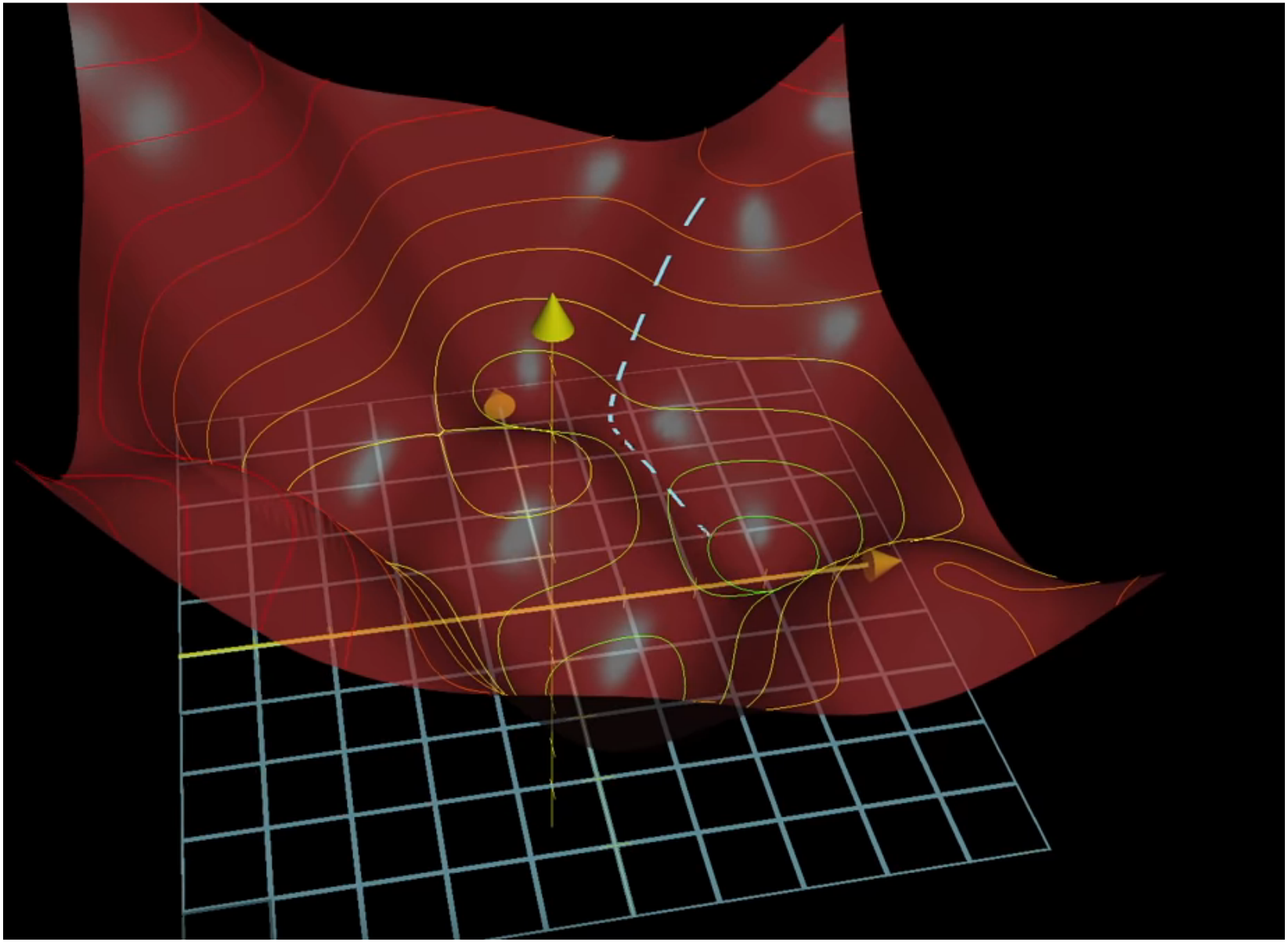
3Blue1Brown

Optymalizacja - Gradient descent



3Blue1Brown

Optymalizacja - Gradient descent



3Blue1Brown