Statystyczna Eksploracja Danych

Wykład 1 - wprowadzenie, metoda LDA Fishera

dr inż. Julian Sienkiewicz

5 marca 2021

Zbieramy i przechowujemy coraz więcej danych:

- jak je wykorzystać?
- jak uzyskać informacje z danych?

Zbieramy i przechowujemy coraz więcej danych:

- jak je wykorzystać?
- jak uzyskać informacje z danych?

Eksploracja danych

Inteligentna analiza danych

Również: *Data Mining*, *Artificial Intelligence*, *Machine learning* (systemy uczące się, sztuczna inteligencja, uczenie maszynowe)

Zbieramy i przechowujemy coraz więcej danych:

- jak je wykorzystać?
- jak uzyskać informacje z danych?

Eksploracja danych

Inteligentna analiza danych

Również: *Data Mining*, *Artificial Intelligence*, *Machine learning* (systemy uczące się, sztuczna inteligencja, uczenie maszynowe)

Definicja wg. Larose'a

Proces odkrywania znaczących nowych powiązań, wzorców i trendów poprzez przeszukiwanie dużych ilości danych zgromadzonych w bazach danych przy użyciu metod matematycznych

Dlaczego teraz?

Dlaczego teraz?

1. wiele cyfrowych czujników np.

Dlaczego teraz?

1. wiele cyfrowych czujników np.
    - sklepy

Dlaczego teraz?

1. wiele cyfrowych czujników np.
   - sklepy
   - kamery na ulicy

Dlaczego teraz?

1. wiele cyfrowych czujników np.
   - sklepy
   - kamery na ulicy
   - GPS w telefonie

Dlaczego teraz?

1. wiele cyfrowych czujników np.
   - sklepy
   - kamery na ulicy
   - GPS w telefonie
2. wiele baz danych: banki, czasopisma, połączenia telefoniczne,

Dlaczego teraz?

1. wiele cyfrowych czujników np.
   - sklepy
   - kamery na ulicy
   - GPS w telefonie
2. wiele baz danych: banki, czasopisma, połączenia telefoniczne,
3. duże moce komputerowe,

Dlaczego teraz?

1. wiele cyfrowych czujników np.
   - sklepy
   - kamery na ulicy
   - GPS w telefonie
2. wiele baz danych: banki, czasopisma, połączenia telefoniczne,
3. duże moce komputerowe,
4. łatwość przesyłu danych.

Dlaczego teraz?

1. wiele cyfrowych czujników np.
   - sklepy
   - kamery na ulicy
   - GPS w telefonie
2. wiele baz danych: banki, czasopisma, połączenia telefoniczne,
3. duże moce komputerowe,
4. łatwość przesyłu danych.

Jak z tego pozyskać użyteczne informacje dla:

Dlaczego teraz?

1. wiele cyfrowych czujników np.
   - sklepy
   - kamery na ulicy
   - GPS w telefonie
2. wiele baz danych: banki, czasopisma, połączenia telefoniczne,
3. duże moce komputerowe,
4. łatwość przesyłu danych.

Jak z tego pozyskać użyteczne informacje dla:

- biznesu, polityki,

Dlaczego teraz?

1. wiele cyfrowych czujników np.
   - sklepy
   - kamery na ulicy
   - GPS w telefonie
2. wiele baz danych: banki, czasopisma, połączenia telefoniczne,
3. duże moce komputerowe,
4. łatwość przesyłu danych.

Jak z tego pozyskać użyteczne informacje dla:

- biznesu, polityki,
- nauki, wojska,

Dlaczego teraz?

1. wiele cyfrowych czujników np.
   - sklepy
   - kamery na ulicy
   - GPS w telefonie
2. wiele baz danych: banki, czasopisma, połączenia telefoniczne,
3. duże moce komputerowe,
4. łatwość przesyłu danych.

Jak z tego pozyskać użyteczne informacje dla:

- biznesu, polityki,
- nauki, wojska,
- sportu, ochrony zdrowia?

Główne zadania eksploracji danych

Główne zadania eksploracji danych

1. opis,

Główne zadania eksploracji danych

1. opis,
2. szacowanie (estymacja),

Główne zadania eksploracji danych

1. opis,
2. szacowanie (estymacja),
3. przewidywanie (predykcja),

Główne zadania eksploracji danych

1. opis,
2. szacowanie (estymacja),
3. przewidywanie (predykcja),
4. klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł,

Główne zadania eksploracji danych

1. opis,
2. szacowanie (estymacja),
3. przewidywanie (predykcja),
4. klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł,
5. grupowanie.

1. Opis

1. Opis

Znaleźć metodę do opisu wzorca lub trendu:

1. Opis

Znaleźć metodę do opisu wzorca lub trendu:

- analizując dane z kolejnych sondaży wyborczych stwierdzamy, że poparcie dla pewnej partii **rośnie** wśród bezrobotnych,

1. Opis

Znaleźć metodę do opisu wzorca lub trendu:

- analizując dane z kolejnych sondaży wyborczych stwierdzamy, że poparcie dla pewnej partii **rośnie** wśród bezrobotnych,
- **nie ma wpływu** bliskość stacji benzynowej na ilość nowotworów wśród mieszkańców osiedla,

1. Opis

Znaleźć metodę do opisu wzorca lub trendu:

- analizując dane z kolejnych sondaży wyborczych stwierdzamy, że poparcie dla pewnej partii **rośnie** wśród bezrobotnych,
- **nie ma wpływu** bliskość stacji benzynowej na ilość nowotworów wśród mieszkańców osiedla,
- student, który zdawał wszystkie egzaminy za pierwszym podejściem **częściej** wybiera specjalność “fizyka komputerowa”,

1. Opis

Znaleźć metodę do opisu wzorca lub trendu:

- analizując dane z kolejnych sondaży wyborczych stwierdzamy, że poparcie dla pewnej partii **rośnie** wśród bezrobotnych,
- **nie ma wpływu** bliskość stacji benzynowej na ilość nowotworów wśród mieszkańców osiedla,
- student, który zdawał wszystkie egzaminy za pierwszym podejściem **częściej** wybiera specjalność “fizyka komputerowa”,
- jakie informacje od farmerów przydają się do **prognozowania** tegorocznych zbiorów określonego rodzaju zboża.


DAS MAGAZIN
LETZTE AUSGABEN
LOGIN
ePAPER
Ich habe nur gezeigt, dass es die Bombe gibt
Der Psychologe Michal Kosinski hat eine Methode entwickelt, um Menschen anhand ihres Verhaltens auf Facebook minutiös zu analysieren. Und verhalf so Donald Trump mit zum Sieg.

Główne zadania eksploracji danych

2. Szacowanie (estymacja)

2. Szacowanie (estymacja)

Szacujemy funkcję zwaną **funkcją celu** na podstawie zmiennych estymacji. Przykłady

2. Szacowanie (estymacja)

Szacujemy funkcję zwaną **funkcją celu** na podstawie zmiennych estymacji. Przykłady

- regresja liniowa (logitowa, kwantylowa),

2. Szacowanie (estymacja)

Szacujemy funkcję zwaną **funkcją celu** na podstawie zmiennych estymacji. Przykłady

- regresja liniowa (logitowa, kwantylowa),
- szacowanie wielkości towaru, który będzie sprzedawany w danym dniu tygodnia w hipermarkecie,

2. Szacowanie (estymacja)

Szacujemy funkcję zwaną **funkcją celu** na podstawie zmiennych estymacji. Przykłady

- regresja liniowa (logitowa, kwantylowa),
- szacowanie wielkości towaru, który będzie sprzedawany w danym dniu tygodnia w hipermarkecie,
- szacowanie liczby minut połączeń telefonicznych dla abonenta określonej grupy,

2. Szacowanie (estymacja)

Szacujemy funkcję zwaną **funkcją celu** na podstawie zmiennych estymacji. Przykłady

- regresja liniowa (logitowa, kwantylowa),
- szacowanie wielkości towaru, który będzie sprzedawany w danym dniu tygodnia w hipermarkecie,
- szacowanie liczby minut połączeń telefonicznych dla abonenta określonej grupy,
- szacowanie spadku ciśnienia tętniczego po podaniu danego leku.

3. Przewidywanie (predykcja)

Przewidywanie

3. Przewidywanie (predykcja)

Przewidywanie

- cen akcji lub kursu waluty w przyszłości,

3. Przewidywanie (predykcja)

Przewidywanie

- cen akcji lub kursu waluty w przyszłości,
- wyników przyszłych wyborów,

3. Przewidywanie (predykcja)

Przewidywanie

- cen akcji lub kursu waluty w przyszłości,
- wyników przyszłych wyborów,
- który zespół wygra mecz,

3. Przewidywanie (predykcja)

Przewidywanie

- cen akcji lub kursu waluty w przyszłości,
- wyników przyszłych wyborów,
- który zespół wygra mecz,
- wysokości nadchodzącej fali powodziowej,

3. Przewidywanie (predykcja)

Przewidywanie

- cen akcji lub kursu waluty w przyszłości,
- wyników przyszłych wyborów,
- który zespół wygra mecz,
- wysokości nadchodzącej fali powodziowej,
- wysokości obrotów firm w nadchodzącym miesiącu (roku),

3. Przewidywanie (predykcja)

Przewidywanie

- cen akcji lub kursu waluty w przyszłości,
- wyników przyszłych wyborów,
- który zespół wygra mecz,
- wysokości nadchodzącej fali powodziowej,
- wysokości obrotów firm w nadchodzącym miesiącu (roku),
- skutków ograniczenia (zwiększenia) prędkości,

3. Przewidywanie (predykcja)

Przewidywanie

- cen akcji lub kursu waluty w przyszłości,
- wyników przyszłych wyborów,
- który zespół wygra mecz,
- wysokości nadchodzącej fali powodziowej,
- wysokości obrotów firm w nadchodzącym miesiącu (roku),
- skutków ograniczenia (zwiększenia) prędkości,
- wielkości inflacji jako skutku zmiany stopy procentowej banku centralnego,

3. Przewidywanie (predykcja)

Przewidywanie

- cen akcji lub kursu waluty w przyszłości,
- wyników przyszłych wyborów,
- który zespół wygra mecz,
- wysokości nadchodzącej fali powodziowej,
- wysokości obrotów firm w nadchodzącym miesiącu (roku),
- skutków ograniczenia (zwiększenia) prędkości,
- wielkości inflacji jako skutku zmiany stopy procentowej banku centralnego,
- wysokości i jakości plonów danej uprawy.

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

Do której grupy zaliczyć (sklasyfikować) dany obiekt?
Wcześniej musimy posiadać **zbiór uczący**. Przykłady:

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

Do której grupy zaliczyć (sklasyfikować) dany obiekt?
Wcześniej musimy posiadać **zbiór uczący**. Przykłady:

- jeżeli klient spełniał nastepujące warunki
  1. był właścicielem firmy,

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

Do której grupy zaliczyć (sklasyfikować) dany obiekt?
Wcześniej musimy posiadać **zbiór uczący**. Przykłady:

- jeżeli klient spełniał nastepujące warunki
  1. był właścicielem firmy,
  2. miał ponad 30 lat,

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

Do której grupy zaliczyć (sklasyfikować) dany obiekt?
Wcześniej musimy posiadać **zbiór uczący**. Przykłady:

- jeżeli klient spełniał nastepujące warunki
  1. był właścicielem firmy,
  2. miał ponad 30 lat,
  3. miał dzieci,

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

Do której grupy zaliczyć (sklasyfikować) dany obiekt?
Wcześniej musimy posiadać **zbiór uczący**. Przykłady:

- jeżeli klient spełniał nastepujące warunki
  1. był właścicielem firmy,
  2. miał ponad 30 lat,
  3. miał dzieci,
  4. wykazywał dochód ponad 20.000 PLN miesięcznie

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

Do której grupy zaliczyć (sklasyfikować) dany obiekt?
Wcześniej musimy posiadać **zbiór uczący**. Przykłady:

- jeżeli klient spełniał nastepujące warunki
  1. był właścicielem firmy,
  2. miał ponad 30 lat,
  3. miał dzieci,
  4. wykazywał dochód ponad 20.000 PLN miesięcznie

  to spłacił pożyczkę,

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

Do której grupy zaliczyć (sklasyfikować) dany obiekt?
Wcześniej musimy posiadać **zbiór uczący**. Przykłady:

- jeżeli klient spełniał nastepujące warunki
  1. był właścicielem firmy,
  2. miał ponad 30 lat,
  3. miał dzieci,
  4. wykazywał dochód ponad 20.000 PLN miesięcznie

  to spłacił pożyczkę,
- jeżeli pacjent ma X lat i ciśnienie tętnicze Y, to użyj leków A, B i C,

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

Do której grupy zaliczyć (sklasyfikować) dany obiekt?
Wcześniej musimy posiadać **zbiór uczący.** Przykłady:

- jeżeli klient spełniał nastepujące warunki
  1. był właścicielem firmy,
  2. miał ponad 30 lat,
  3. miał dzieci,
  4. wykazywał dochód ponad 20.000 PLN miesięcznie

  to spłacił pożyczkę,
- jeżeli pacjent ma X lat i ciśnienie tętnicze Y, to użyj leków A, B i C,
- jeżeli e-mail zawiera słowa

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

Do której grupy zaliczyć (sklasyfikować) dany obiekt?
Wcześniej musimy posiadać **zbiór uczący.** Przykłady:

- jeżeli klient spełniał nastepujące warunki
  1. był właścicielem firmy,
  2. miał ponad 30 lat,
  3. miał dzieci,
  4. wykazywał dochód ponad 20.000 PLN miesięcznie

  to spłacił pożyczkę,
- jeżeli pacjent ma X lat i ciśnienie tętnicze Y, to użyj leków A, B i C,
- jeżeli e-mail zawiera słowa
  1. SEX

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

Do której grupy zaliczyć (sklasyfikować) dany obiekt?
Wcześniej musimy posiadać **zbiór uczący**. Przykłady:

- jeżeli klient spełniał nastepujące warunki
  1. był właścicielem firmy,
  2. miał ponad 30 lat,
  3. miał dzieci,
  4. wykazywał dochód ponad 20.000 PLN miesięcznie

  to spłacił pożyczkę,
- jeżeli pacjent ma X lat i ciśnienie tętnicze Y, to użyj leków A, B i C,
- jeżeli e-mail zawiera słowa
  1. SEX
  2. YOU WON LOTTERY

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

Do której grupy zaliczyć (sklasyfikować) dany obiekt?
Wcześniej musimy posiadać **zbiór uczący**. Przykłady:

- jeżeli klient spełniał nastepujące warunki
  1. był właścicielem firmy,
  2. miał ponad 30 lat,
  3. miał dzieci,
  4. wykazywał dochód ponad 20.000 PLN miesięcznie

  to spłacił pożyczkę,
- jeżeli pacjent ma X lat i ciśnienie tętnicze Y, to użyj leków A, B i C,
- jeżeli e-mail zawiera słowa
  1. SEX
  2. YOU WON LOTTERY
  3. PASSWORD REQUEST

4. Klasyfikacja (uczenie pod nadzorem) + odkrywanie reguł

Do której grupy zaliczyć (sklasyfikować) dany obiekt?
Wcześniej musimy posiadać **zbiór uczący**. Przykłady:

- jeżeli klient spełniał nastepujące warunki
  1. był właścicielem firmy,
  2. miał ponad 30 lat,
  3. miał dzieci,
  4. wykazywał dochód ponad 20.000 PLN miesięcznie

  to spłacił pożyczkę,
- jeżeli pacjent ma X lat i ciśnienie tętnicze Y, to użyj leków A, B i C,
- jeżeli e-mail zawiera słowa
  1. SEX
  2. YOU WON LOTTERY
  3. PASSWORD REQUEST

  przenieś go do folderu SPAM.

5. Grupowanie

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

- samochody firmy X model Y kupują

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

- samochody firmy X model Y kupują
  - ambasadorowie krajów Trzeciego Świata,

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

- samochody firmy X model Y kupują
  - ambasadorowie krajów Trzeciego Świata,
  - biznesmeni z dochodem rocznym 100.000-200.000 dolarów,

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

- samochody firmy X model Y kupują
  - ambasadorowie krajów Trzeciego Świata,
  - biznesmeni z dochodem rocznym 100.000-200.000 dolarów,
  - artyści jazzowi w wieku 50-60 lat

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

- samochody firmy X model Y kupują
  - ambasadorowie krajów Trzeciego Świata,
  - biznesmeni z dochodem rocznym 100.000-200.000 dolarów,
  - artyści jazzowi w wieku 50-60 lat

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

- samochody firmy X model Y kupują
  - ambasadorowie krajów Trzeciego Świata,
  - biznesmeni z dochodem rocznym 100.000-200.000 dolarów,
  - artyści jazzowi w wieku 50-60 lat
- zmiany cen akcji firmy X to głównie

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

- samochody firmy X model Y kupują
  - ambasadorowie krajów Trzeciego Świata,
  - biznesmeni z dochodem rocznym 100.000-200.000 dolarów,
  - artyści jazzowi w wieku 50-60 lat
- zmiany cen akcji firmy X to głównie
  - spadki o 0.8–1.2 %,

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

- samochody firmy X model Y kupują
    - ambasadorowie krajów Trzeciego Świata,
    - biznesmeni z dochodem rocznym 100.000-200.000 dolarów,
    - artyści jazzowi w wieku 50-60 lat
- zmiany cen akcji firmy X to głównie
    - spadki o 0.8–1.2 %,
    - wzrosty o 0.9–1.3 %

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

- samochody firmy X model Y kupują
  - ambasadorowie krajów Trzeciego Świata,
  - biznesmeni z dochodem rocznym 100.000-200.000 dolarów,
  - artyści jazzowi w wieku 50-60 lat
- zmiany cen akcji firmy X to głównie
  - spadki o 0.8–1.2 %,
  - wzrosty o 0.9–1.3 %

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

- samochody firmy X model Y kupują
  - ambasadorowie krajów Trzeciego Świata,
  - biznesmeni z dochodem rocznym 100.000-200.000 dolarów,
  - artyści jazzowi w wieku 50-60 lat
- zmiany cen akcji firmy X to głównie
  - spadki o 0.8–1.2 %,
  - wzrosty o 0.9–1.3 %
- chorzy w okresie listopad-grudzień cierpią głównie na

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

- samochody firmy X model Y kupują
  - ambasadorowie krajów Trzeciego Świata,
  - biznesmeni z dochodem rocznym 100.000-200.000 dolarów,
  - artyści jazzowi w wieku 50-60 lat
- zmiany cen akcji firmy X to głównie
  - spadki o 0.8–1.2 %,
  - wzrosty o 0.9–1.3 %
- chorzy w okresie listopad-grudzień cierpią głównie na
  - grypę,

5. Grupowanie

Algorytm próbuje podzielić wszystkie dane na kilka wewnętrznie podobnych grup, nie wiedząc, jakie są kryteria produktu ani też jakie są grupy (analiza skupień). Przykłady:

- samochody firmy X model Y kupują
  - ambasadorowie krajów Trzeciego Świata,
  - biznesmeni z dochodem rocznym 100.000-200.000 dolarów,
  - artyści jazzowi w wieku 50-60 lat
- zmiany cen akcji firmy X to głównie
  - spadki o 0.8–1.2 %,
  - wzrosty o 0.9–1.3 %
- chorzy w okresie listopad-grudzień cierpią głównie na
  - grypę,
  - przeziębienia.

Liniowa analiza dyskryminacji

Linear discriminant analysis (LDA) stara się zredukować wymiarowość problemu, zachowując tak wiele informacji o pierwotnym zbiorze, jak tylko można. Metoda, stworzona przez **sir Ronalda A. Fishera**, polega na rzutowaniu obserwacji na optymalny kierunek w przestrzeni.

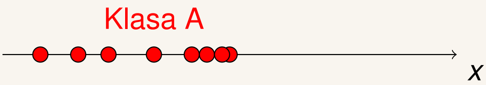
Ronald Fisher

- 1890 — 1962,
- genetyk i statystyk brytyjski,
- twórca takich pojęć jak metoda największej wiarygodności (ang. *maximum likelihood*), analiza wariancji (*ANOVA*), test Fishera (*F-test*) czy informacja Fishera.

W skrócie: jak odróżnić, do której klasy należy dany przypadek?

Przypadek jednowymiarowy

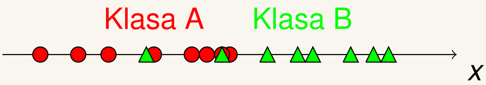
Mamy dwie klasy: **A** i **B**. Każda obserwacja *i* ma wartość x_i.



W skrócie: jak odróżnić, do której klasy należy dany przypadek?

Przypadek jednowymiarowy

Mamy dwie klasy: **A** i **B**. Każda obserwacja *i* ma wartość x_i.

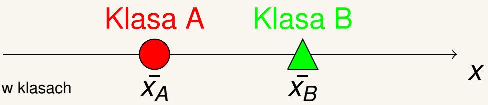


W skrócie: jak odróżnić, do której klasy należy dany przypadek?

Przypadek jednowymiarowy

Mamy dwie klasy: **A** i **B**. Każda obserwacja i ma wartość x_i.

Liczymy średnią arytmetyczną w klasach

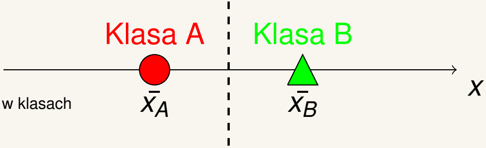


W skrócie: jak odróżnić, do której klasy należy dany przypadek?

Przypadek jednowymiarowy

Mamy dwie klasy: **A** i **B**. Każda obserwacja *i* ma wartość x_i.

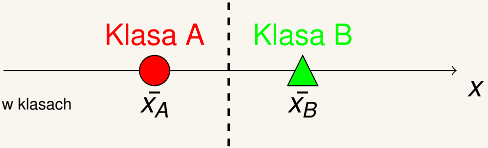
Liczymy średnią arytmetyczną w klasach



W skrócie: jak odróżnić, do której klasy należy dany przypadek?

Przypadek jednowymiarowy

Mamy dwie klasy: **A** i **B**. Każda obserwacja i ma wartość x_i.



Liczymy średnią arytmetyczną w klasach

Liczymy średnią arytmetyczną $x^* = \frac{1}{2}\left(\bar{x}_A + \bar{x}_B\right)$

Reguła decyzyjna

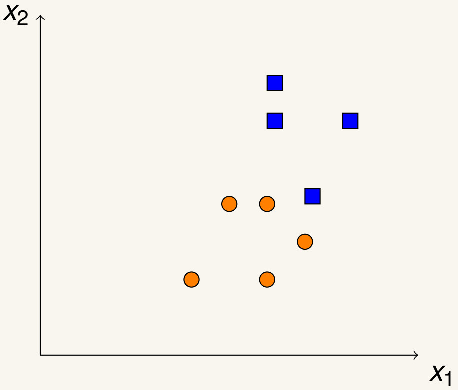
Pojawia się nowa obserwacja j. Do której klasy ją zaliczymy?

Jezeli $\begin{cases} x_j < x^* & \Longrightarrow \ x_j \in \mathcal{A} \\ x_j > x^* & \Longrightarrow \ x_j \in \mathcal{B} \end{cases}$

A co w dwóch wymiarach?

Przypadek dwuwymiarowy

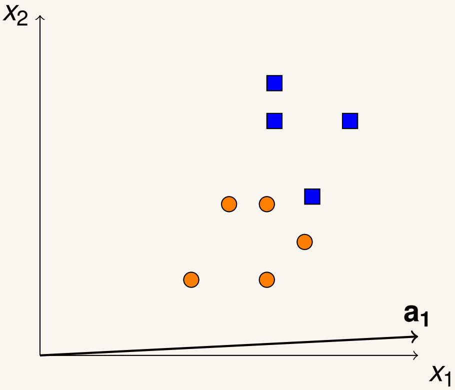
Mamy dwie klasy: **A** i **B**. Szukamy kierunku **a**, na który można zrzutować obserwacje.



A co w dwóch wymiarach?

Przypadek dwuwymiarowy

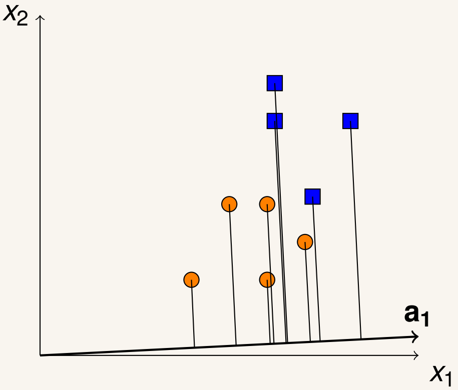
Mamy dwie klasy: **A** i **B**. Szukamy kierunku **a**, na który można zrzutować obserwacje.



A co w dwóch wymiarach?

Przypadek dwuwymiarowy

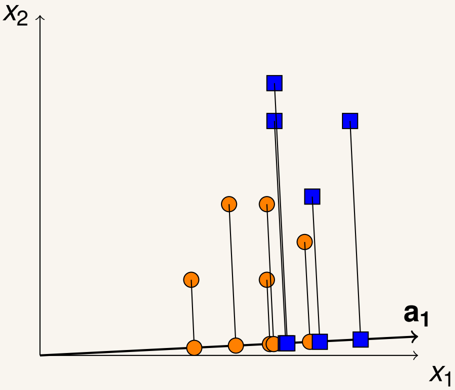
Mamy dwie klasy: **A** i **B**. Szukamy kierunku **a**, na który można zrzutować obserwacje.



A co w dwóch wymiarach?

Przypadek dwuwymiarowy

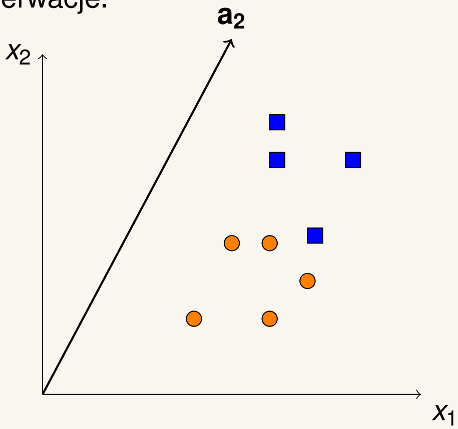
Mamy dwie klasy: **A** i **B**. Szukamy kierunku **a**, na który można zrzutować obserwacje.



A co w dwóch wymiarach?

Przypadek dwuwymiarowy

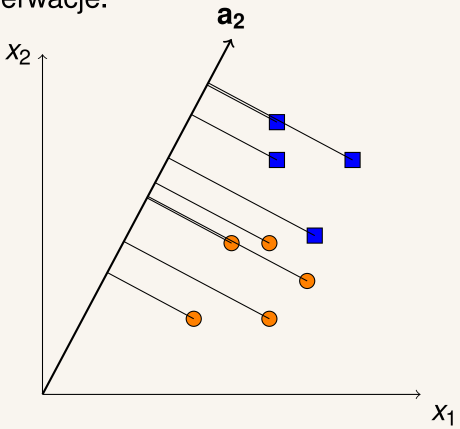
Mamy dwie klasy: **A** i **B**. Szukamy kierunku **a**, na który można zrzutować obserwacje.



A co w dwóch wymiarach?

Przypadek dwuwymiarowy

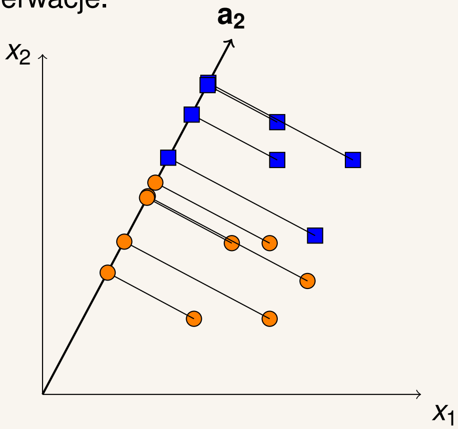
Mamy dwie klasy: **A** i **B**. Szukamy kierunku **a**, na który można zrzutować obserwacje.



A co w dwóch wymiarach?

Przypadek dwuwymiarowy

Mamy dwie klasy: **A** i **B**. Szukamy kierunku **a**, na który można zrzutować obserwacje.



Mamy próby należące do dwóch klas $y \in \{A, B\}$: $(\mathbf{x_1}, y_1), (\mathbf{x_2}, y_2), ...,$ $(\mathbf{x_n}, y_n)$, gdzie $\mathbf{x_i}$ to i-ta obserwacja, a y_i to jej przynależność do jednej z klas.

Każdy element $\mathbf{x}$ jest wektorem **kolumnowym** o długości p

$$\bar{\mathbf{x}}_\mathbf{i} = \begin{bmatrix} x_i^1 \\ x_i^2 \\ \dots \\ x_i^p \end{bmatrix}$$

Wartość p określa wymiar układu (ogólnie niekoniecznie przestrzeń fizyczna).

W kadej klasie możemy wyznaczyć wartość średnią

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{i=n} \mathbf{x_i}$$

oraz macierz kowariancji

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{i=n} (\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})^T$$

Główne założenie metody LDA: macierze kowariancji **w obu klasach** są takie same. W efekcie macierz kowariancji wewnątrzgrupowej **W** wyraża się wzorem:

$$\mathbf{W} = \frac{1}{n-2}\sum_{k=1}^{2}(n_k - 1)\mathbf{S_k} = \frac{1}{n-2}\sum_{k=1}^{2}\sum_{i=1}^{n_k}(\mathbf{x_{ki}} - \bar{\mathbf{x}}_k)(\mathbf{x_{ki}} - \bar{\mathbf{x}}_k)^T$$

$\mathbf{S_k}$ - macierz kowariancji w klasie k,
$n = n_1 + n_2$ - liczba danych w obu klasach

Oczywiście, w rzeczywistości ciężko znaleźć przypadki o takich samych macierzach kowariancji, ale podejście dość dobrze pracuje nawet w przypadku odchyłek.

Reguła dyskryminacyjna

Znajdź taki kierunek $\widetilde{\mathbf{a}}$, który najlepiej rozdziela podpróby uczące. Za miarę rozdzielności weź kwadrat odległości pomiędzy średnimi arytmetycznymi wzdłuż kierunku $\mathbf{a}$, znormalizowany przez zmienność klas w kierunku $\mathbf{a}$.

$$J = \frac{\left(\mathbf{a}^T\bar{\mathbf{x}}_2 - \mathbf{a}^T\bar{\mathbf{x}}_1\right)^2}{\mathbf{a}^T\mathbf{W}\mathbf{a}} \tag{1}$$

Zmienność obserwacji wewnątrz klas w kierunku $\mathbf{a}$:

$$Var(\mathbf{a}^T\mathbf{x}) = \mathbf{a}^T\mathbf{W}\,\mathbf{a}$$

Czyli szukamy wektora maksymalizującego wyrażenie dane równaniem (1). W tym celu różniczkujemy (1) po **a**

$$\frac{dJ}{d\mathbf{a}} = \frac{d}{d\mathbf{a}}\left[\frac{\left(\mathbf{a}^T\bar{\mathbf{x}}_2 - \mathbf{a}^T\bar{\mathbf{x}}_1\right)^2}{\mathbf{a}^T\mathbf{W}\mathbf{a}}\right] = \tag{2}$$

Czyli szukamy wektora maksymalizującego wyrażenie dane równaniem (1). W tym celu różniczkujemy (1) po **a**

$$\frac{dJ}{d\mathbf{a}} = \frac{d}{d\mathbf{a}}\left[\frac{\left(\mathbf{a}^T\bar{\mathbf{x}}_2 - \mathbf{a}^T\bar{\mathbf{x}}_1\right)^2}{\mathbf{a}^T\mathbf{W}\mathbf{a}}\right] = \tag{2}$$

$$= \frac{2\mathbf{a}^T\mathbf{W}\mathbf{a}\mathbf{a}^T(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^T - 2[\mathbf{a}^T(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)]^2\mathbf{a}^T\mathbf{W}}{(\mathbf{a}^T\mathbf{W}\mathbf{a})^2}$$

Czyli szukamy wektora maksymalizującego wyrażenie dane równaniem (1). W tym celu różniczkujemy (1) po **a**

$$\frac{dJ}{d\mathbf{a}} = \frac{d}{d\mathbf{a}}\left[\frac{\left(\mathbf{a}^T\bar{\mathbf{x}}_2 - \mathbf{a}^T\bar{\mathbf{x}}_1\right)^2}{\mathbf{a}^T\mathbf{W}\mathbf{a}}\right] = \tag{2}$$

$$= \frac{2\mathbf{a}^T\mathbf{W}\mathbf{a}\mathbf{a}^T(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^T - 2[\mathbf{a}^T(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)]^2\mathbf{a}^T\mathbf{W}}{(\mathbf{a}^T\mathbf{W}\mathbf{a})^2}$$

i przyrównujemy mianownik do zera. Wyróżnione elementy są skalarami, oznaczymy je jako A i B:

Czyli szukamy wektora maksymalizującego wyrażenie dane równaniem (1). W tym celu różniczkujemy (1) po **a**

$$\frac{dJ}{d\mathbf{a}} = \frac{d}{d\mathbf{a}}\left[\frac{\left(\mathbf{a}^T\bar{\mathbf{x}}_2 - \mathbf{a}^T\bar{\mathbf{x}}_1\right)^2}{\mathbf{a}^T\mathbf{W}\mathbf{a}}\right] = \tag{2}$$

$$= \frac{2\mathbf{a}^T\mathbf{W}\mathbf{a}\mathbf{a}^T(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^T - 2[\mathbf{a}^T(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)]^2\mathbf{a}^T\mathbf{W}}{(\mathbf{a}^T\mathbf{W}\mathbf{a})^2}$$

i przyrównujemy mianownik do zera. Wyróżnione elementy są skalarami, oznaczymy je jako A i B:

$$A\left(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1\right)^T - B\,\mathbf{a}^T\mathbf{W} = 0$$

Ostatecznie, wektor $\widetilde{\mathbf{a}}$, maksymalizujący wyrażenie J ma postać

(Pierwszy) wektor kanoniczny

$$\widetilde{\mathbf{a}} \sim \mathbf{W}^{-1}(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)$$

Ostatecznie, wektor $\widetilde{\mathbf{a}}$, maksymalizujący wyrażenie J ma postać

(Pierwszy) wektor kanoniczny

$$\widetilde{\mathbf{a}} \sim \mathbf{W}^{-1}(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)$$

Czyli obserwacja $\mathbf{x}$ spełniająca warunek

Warunek dyskryminacyjny

$$|\widetilde{\mathbf{a}}^T(\mathbf{x} - \bar{\mathbf{x}}_1)| < |\widetilde{\mathbf{a}}^T(\mathbf{x} - \bar{\mathbf{x}}_2)|$$

zostanie uznana za przynależną do klasy A.

Ostatecznie, wektor $\widetilde{\mathbf{a}}$, maksymalizujący wyrażenie J ma postać

(Pierwszy) wektor kanoniczny

$$\widetilde{\mathbf{a}} \sim \mathbf{W}^{-1}(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)$$

Czyli obserwacja $\mathbf{x}$ spełniająca warunek

Warunek dyskryminacyjny

$$|\widetilde{\mathbf{a}}^T(\mathbf{x} - \bar{\mathbf{x}}_1)| < |\widetilde{\mathbf{a}}^T(\mathbf{x} - \bar{\mathbf{x}}_2)|$$

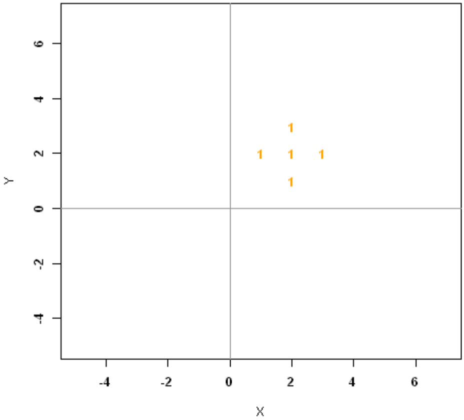
zostanie uznana za przynależną do klasy A. Natomiast hiperpłaszczyzna dyskryminacyjna jest dana równaniem

Hiperpłaszczyzna dyskryminacyjna

$$(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^T\mathbf{W}^{-1}\left[\mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)\right] = 0$$

Przykład 1 - 2D (rozkłady symetryczne)

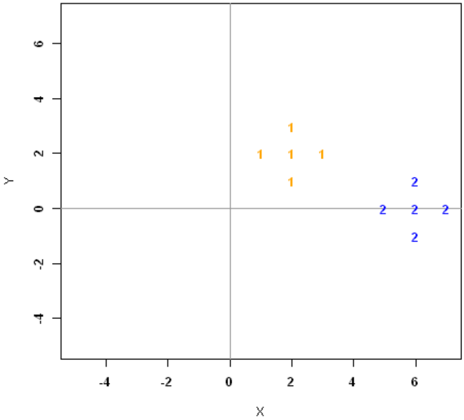
$$\mathbf{x_1} = \begin{bmatrix} 2 & 2 & 2 & 1 & 3 \\ 2 & 1 & 3 & 2 & 2 \end{bmatrix}$$



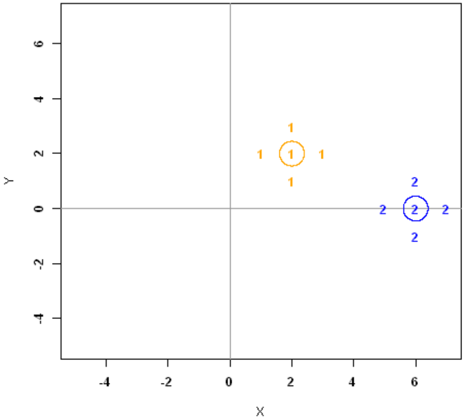
Przykład 1 - 2D (rozkłady symetryczne)

$$\mathbf{x_1} = \begin{bmatrix} 2 & 2 & 2 & 1 & 3 \\ 2 & 1 & 3 & 2 & 2 \end{bmatrix}$$

$$\mathbf{x_2} = \begin{bmatrix} 6 & 6 & 6 & 5 & 7 \\ 0 & -1 & 1 & 0 & 0 \end{bmatrix}$$



Przykład 1 - 2D (rozkłady symetryczne)



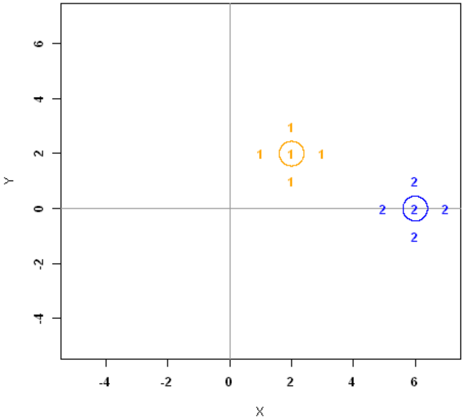
$$\mathbf{x_1} = \begin{bmatrix} 2 & 2 & 2 & 1 & 3 \\ 2 & 1 & 3 & 2 & 2 \end{bmatrix}$$

$$\mathbf{x_2} = \begin{bmatrix} 6 & 6 & 6 & 5 & 7 \\ 0 & -1 & 1 & 0 & 0 \end{bmatrix}$$

$$\bar{\mathbf{x_1}} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \mathbf{S_1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\bar{\mathbf{x_2}} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \mathbf{S_2} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

Przykład 1 - 2D
(rozkłady symetryczne)



$$\mathbf{x_1} = \begin{bmatrix} 2 & 2 & 2 & 1 & 3 \\ 2 & 1 & 3 & 2 & 2 \end{bmatrix}$$

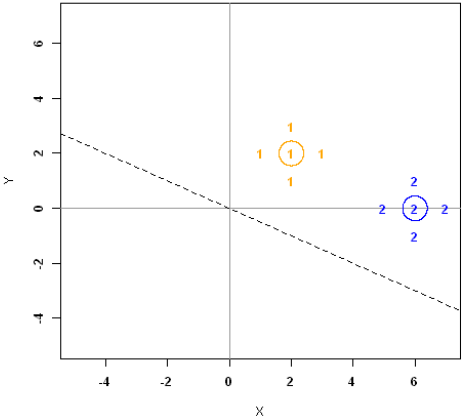
$$\mathbf{x_2} = \begin{bmatrix} 6 & 6 & 6 & 5 & 7 \\ 0 & -1 & 1 & 0 & 0 \end{bmatrix}$$

$$\bar{\mathbf{x_1}} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \mathbf{S_1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\bar{\mathbf{x_2}} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \quad \mathbf{S_2} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \quad \mathbf{W}^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Przykład 1 - 2D (rozkłady symetryczne)



$$\mathbf{x_1} = \begin{bmatrix} 2 & 2 & 2 & 1 & 3 \\ 2 & 1 & 3 & 2 & 2 \end{bmatrix}$$

$$\mathbf{x_2} = \begin{bmatrix} 6 & 6 & 6 & 5 & 7 \\ 0 & -1 & 1 & 0 & 0 \end{bmatrix}$$

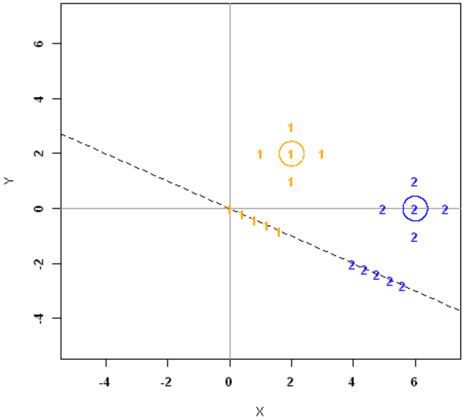
$$\bar{\mathbf{x_1}} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \mathbf{S_1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\bar{\mathbf{x_2}} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \quad \mathbf{S_2} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \quad \mathbf{W}^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\mathbf{a} \sim \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 8 \\ -4 \end{bmatrix}$$

Przykład 1 - 2D
(rozkłady symetryczne)



$$\mathbf{x_1} = \begin{bmatrix} 2 & 2 & 2 & 1 & 3 \\ 2 & 1 & 3 & 2 & 2 \end{bmatrix}$$

$$\mathbf{x_2} = \begin{bmatrix} 6 & 6 & 6 & 5 & 7 \\ 0 & -1 & 1 & 0 & 0 \end{bmatrix}$$

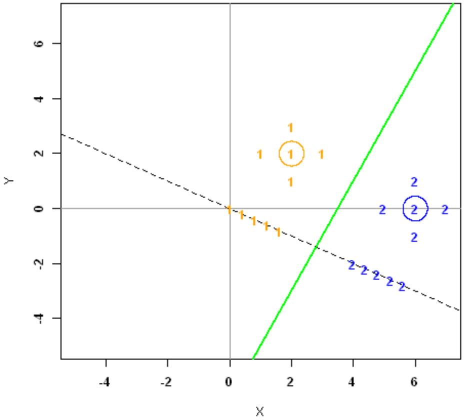
$$\bar{\mathbf{x_1}} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \; \mathbf{S_1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\bar{\mathbf{x_2}} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \; \mathbf{S_2} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \; \mathbf{W}^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\mathbf{a} \sim \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 8 \\ -4 \end{bmatrix}$$

Przykład 1 - 2D (rozkłady symetryczne)



$$\mathbf{x_1} = \begin{bmatrix} 2 & 2 & 2 & 1 & 3 \\ 2 & 1 & 3 & 2 & 2 \end{bmatrix}$$

$$\mathbf{x_2} = \begin{bmatrix} 6 & 6 & 6 & 5 & 7 \\ 0 & -1 & 1 & 0 & 0 \end{bmatrix}$$

$$\bar{\mathbf{x_1}} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \mathbf{S_1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

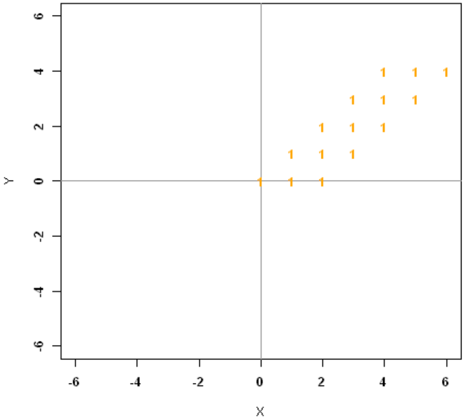
$$\bar{\mathbf{x_2}} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \mathbf{S_2} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \mathbf{W}^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

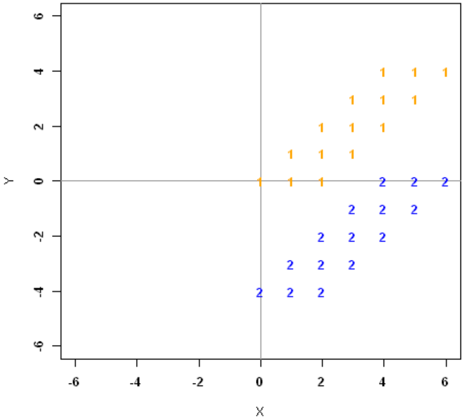
$$\mathbf{a} \sim \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 8 \\ -4 \end{bmatrix}$$

$$y = 2x - 7$$

Przykład 2 - 2D (rozkłady eliptyczne)



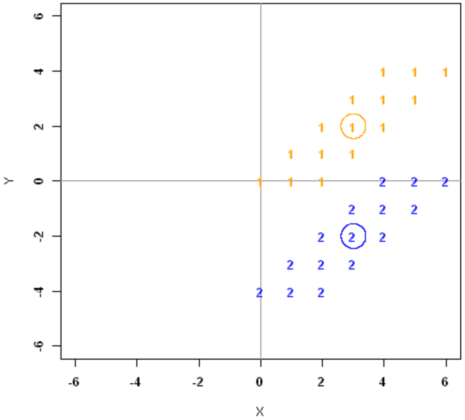
Przykład 2 - 2D (rozkłady eliptyczne)



Przykład 2 - 2D
(rozkłady eliptyczne)

$$\bar{x}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \; \mathbf{S}_1 = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

$$\bar{x}_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \; \mathbf{S}_2 = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$



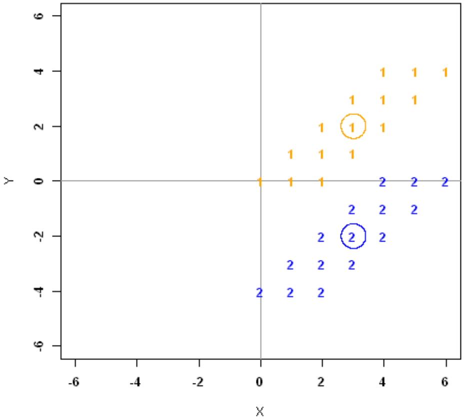
Przykład 2 - 2D (rozkłady eliptyczne)

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad \mathbf{S}_1 = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

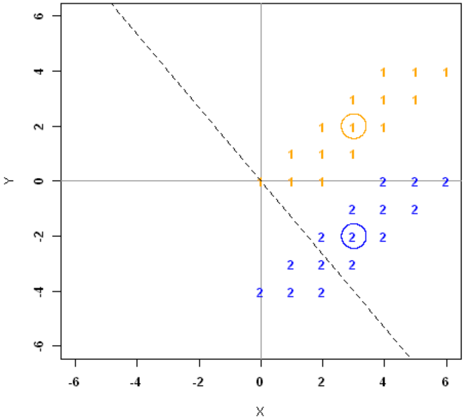
$$\bar{\mathbf{x}}_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \quad \mathbf{S}_2 = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

$$\mathbf{W}^{-1} = \begin{bmatrix} \frac{7}{5} & -\frac{7}{5} \\ -\frac{7}{5} & \frac{28}{15} \end{bmatrix}$$



Przykład 2 - 2D (rozkłady eliptyczne)



$$\bar{x}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad \mathbf{S_1} = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

$$\bar{x}_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \quad \mathbf{S_2} = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

$$\mathbf{W}^{-1} = \begin{bmatrix} \frac{7}{5} & -\frac{7}{5} \\ -\frac{7}{5} & \frac{28}{15} \end{bmatrix}$$

$$\mathbf{a} \sim \begin{bmatrix} \frac{28}{5} \\ -\frac{112}{15} \end{bmatrix} \sim \begin{bmatrix} 1 \\ -1\frac{1}{3} \end{bmatrix}$$

Przykład 2 - 2D
(rozkłady eliptyczne)

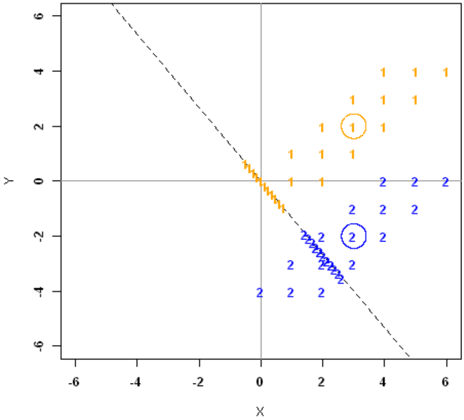
$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad \mathbf{S_1} = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

$$\bar{\mathbf{x}}_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \quad \mathbf{S_2} = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

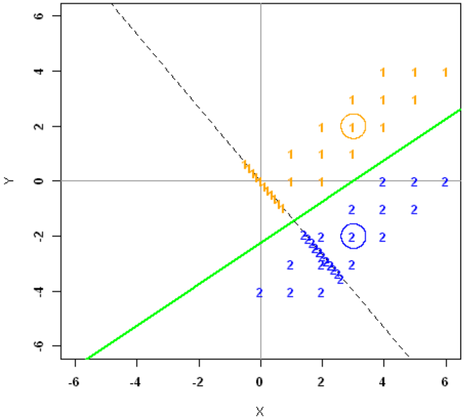
$$\mathbf{W} = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

$$\mathbf{W}^{-1} = \begin{bmatrix} \frac{7}{5} & -\frac{7}{5} \\ -\frac{7}{5} & \frac{28}{15} \end{bmatrix}$$

$$\mathbf{a} \sim \begin{bmatrix} \frac{28}{5} \\ -\frac{112}{15} \end{bmatrix} \sim \begin{bmatrix} 1 \\ -1\frac{1}{3} \end{bmatrix}$$



Przykład 2 - 2D (rozkłady eliptyczne)



$$\bar{x_1} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad \mathbf{S_1} = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

$$\bar{x_2} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \quad \mathbf{S_2} = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

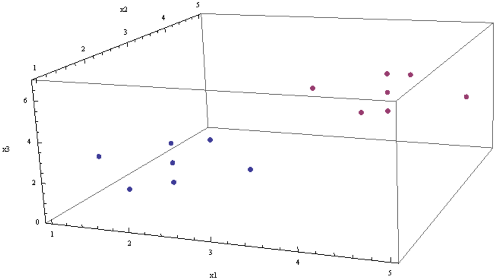
$$\mathbf{W} = \begin{bmatrix} \frac{20}{7} & \frac{15}{7} \\ \frac{15}{7} & \frac{15}{7} \end{bmatrix}$$

$$\mathbf{W}^{-1} = \begin{bmatrix} \frac{7}{5} & -\frac{7}{5} \\ -\frac{7}{5} & \frac{28}{15} \end{bmatrix}$$

$$\mathbf{a} \sim \begin{bmatrix} \frac{28}{5} \\ -\frac{112}{15} \end{bmatrix} \sim \begin{bmatrix} 1 \\ -1\frac{1}{3} \end{bmatrix}$$

$$y = x - 3$$

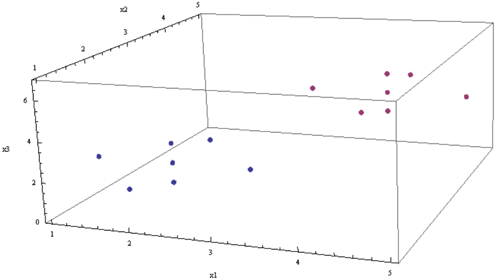
Przykład 3 - 3D



$$\mathbf{x_1} = \begin{bmatrix} 2 & 2 & 2 & 1 & 3 & 2 & 2 \\ 2 & 1 & 3 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 & 1 & 3 \end{bmatrix}$$

$$\mathbf{x_2} = \begin{bmatrix} 4 & 4 & 4 & 3 & 5 & 4 & 4 \\ 4 & 3 & 5 & 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & 4 & 4 & 3 & 5 \end{bmatrix}$$

Przykład 3 - 3D



$$\mathbf{x_1} = \begin{bmatrix} 2 & 2 & 2 & 1 & 3 & 2 & 2 \\ 2 & 1 & 3 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 & 1 & 3 \end{bmatrix}$$

$$\mathbf{x_2} = \begin{bmatrix} 4 & 4 & 4 & 3 & 5 & 4 & 4 \\ 4 & 3 & 5 & 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & 4 & 4 & 3 & 5 \end{bmatrix}$$

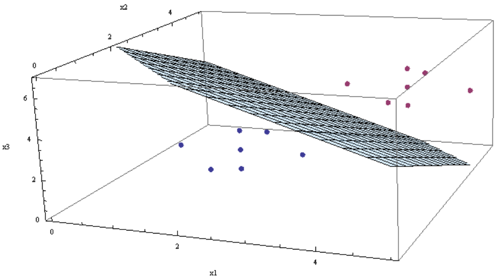
$$\bar{\mathbf{x_1}} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \quad \mathbf{S_1} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

$$\bar{\mathbf{x_2}} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} \quad \mathbf{S_2} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

$$\mathbf{W}^{-1} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Przykład 3 - 3D



$$\mathbf{x_1} = \begin{bmatrix} 2 & 2 & 2 & 1 & 3 & 2 & 2 \\ 2 & 1 & 3 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 & 1 & 3 \end{bmatrix}$$

$$\mathbf{x_2} = \begin{bmatrix} 4 & 4 & 4 & 3 & 5 & 4 & 4 \\ 4 & 3 & 5 & 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & 4 & 4 & 3 & 5 \end{bmatrix}$$

$$\bar{\mathbf{x_1}} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \quad \mathbf{S_1} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

$$\bar{\mathbf{x_2}} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} \quad \mathbf{S_2} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

$$\mathbf{W}^{-1} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$\mathbf{a} \sim \begin{bmatrix} 6 \\ 6 \\ 6 \end{bmatrix} \quad z = -x - y - 9$$

Można zauważyć, że

$$\left(\mathbf{a}^T\bar{\mathbf{x}}_\mathbf{2} - \mathbf{a}^T\bar{\mathbf{x}}_\mathbf{1}\right)^2 = \mathbf{a}^T(\bar{\mathbf{x}}_\mathbf{2} - \bar{\mathbf{x}}_\mathbf{1})(\bar{\mathbf{x}}_\mathbf{2} - \bar{\mathbf{x}}_\mathbf{1})^T\mathbf{a} = \tag{3}$$
$$= \frac{n_1 + n_2}{n_1 n_2}\mathbf{a}^T\sum_{k=1}^{2} n_k(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})^T\mathbf{a}$$

gdzie $\bar{\mathbf{x}}$ to średnia ogólna ze wszystkich obserwacji.

Można zauważyć, że

$$\left(\mathbf{a}^T\bar{\mathbf{x}}_\mathbf{2} - \mathbf{a}^T\bar{\mathbf{x}}_\mathbf{1}\right)^2 = \mathbf{a}^T(\bar{\mathbf{x}}_\mathbf{2} - \bar{\mathbf{x}}_\mathbf{1})(\bar{\mathbf{x}}_\mathbf{2} - \bar{\mathbf{x}}_\mathbf{1})^T\mathbf{a} = \\ = \frac{n_1 + n_2}{n_1 n_2}\mathbf{a}^T\sum_{k=1}^{2} n_k(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})^T\mathbf{a} \tag{3}$$

gdzie $\bar{\mathbf{x}}$ to średnia ogólna ze wszystkich obserwacji. Możemy dzięki temu zdefiniować macierz charakteryzującą **zmienność międzygrupową**

Macierz kowariancji międzygrupowej

$$\mathbf{B} = \sum_{k=1}^{2} n_k(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})^T$$

W efekcie maksymalizacja ze względu na **a** ilorazu

$$\frac{\left(\mathbf{a}^T\bar{\mathbf{x}}_2 - \mathbf{a}^T\bar{\mathbf{x}}_1\right)^2}{\mathbf{a}^T\mathbf{W}\mathbf{a}}$$

W efekcie maksymalizacja ze względu na **a** ilorazu

$$\frac{\left(\mathbf{a}^T\bar{\mathbf{x}}_2 - \mathbf{a}^T\bar{\mathbf{x}}_1\right)^2}{\mathbf{a}^T\mathbf{W}\mathbf{a}}$$

jest równoznaczna maksymalizacji wyrażenia

$$\frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{W}\mathbf{a}},$$

W efekcie maksymalizacja ze względu na **a** ilorazu

$$\frac{\left(\mathbf{a}^T\bar{\mathbf{x}}_2 - \mathbf{a}^T\bar{\mathbf{x}}_1\right)^2}{\mathbf{a}^T\mathbf{W}\mathbf{a}}$$

jest równoznaczna maksymalizacji wyrażenia

$$\frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{W}\mathbf{a}},$$

co daje możliwość uogólnienia metody Fishera na liczbę $g > 2$ klas.

Problem dyskryminacyjny $g > 2$

Znajdź kierunek $\widetilde{\mathbf{a}}$ maksymalizujący wyrażenie

$$\frac{\mathbf{a}^T \mathbf{B} \mathbf{a}}{\mathbf{a}^T \mathbf{W} \mathbf{a}}$$

Problem dyskryminacyjny $g > 2$

Znajdź kierunek $\widetilde{\mathbf{a}}$ maksymalizujący wyrażenie

$$\frac{\mathbf{a}^T \mathbf{B} \mathbf{a}}{\mathbf{a}^T \mathbf{W} \mathbf{a}}$$

gdzie

Macierz kowariancji międzygrupowej

$$\mathbf{B} = \frac{1}{g-1} \sum_{k=1}^{g} n_k (\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})^T$$

Problem dyskryminacyjny $g > 2$

Znajdź kierunek $\widetilde{\mathbf{a}}$ maksymalizujący wyrażenie

$$\frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{W}\mathbf{a}}$$

gdzie

Macierz kowariancji międzygrupowej

$$\mathbf{B} = \frac{1}{g-1}\sum_{k=1}^{g} n_k(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})^T$$

i

Macierz kowariancji wewnątrzgrupowej

$$\mathbf{W} = \frac{1}{n-g}\sum_{k=1}^{g} n_k(n_k - 1)\mathbf{S}_k$$

Rozwiązanie $g > 2$

Wektor $\widetilde{\mathbf{a}}$ maksymalizujący wyrażenie

$$\frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{W}\mathbf{a}}$$

jest **wektorem własnym** macierzy $\mathbf{W}^{-1}\mathbf{B}$, odpowiadającym największej **wartości własnej** tej macierzy

Rozwiązanie $g > 2$

Wektor $\widetilde{\mathbf{a}}$ maksymalizujący wyrażenie

$$\frac{\mathbf{a}^T \mathbf{B} \mathbf{a}}{\mathbf{a}^T \mathbf{W} \mathbf{a}}$$

jest **wektorem własnym** macierzy $\mathbf{W}^{-1}\mathbf{B}$, odpowiadającym największej **wartości własnej** tej macierzy

Reguła decyzyjna

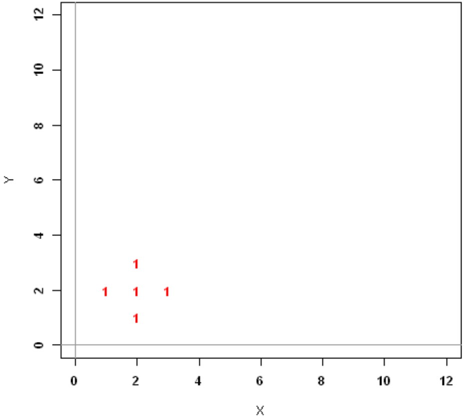
Obserwację $\mathbf{x}$ przypisujemy do klasy j ,jeżeli

$$|\widetilde{\mathbf{a}}^T(\mathbf{x} - \bar{\mathbf{x}}_\mathbf{j})| < |\widetilde{\mathbf{a}}^T(\mathbf{x} - \bar{\mathbf{x}}_\mathbf{k})|$$

dla wszystkich $k \neq j$.

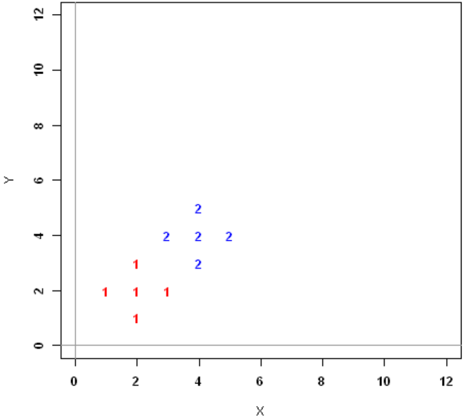
Przykład 4 - $g = 5$

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$



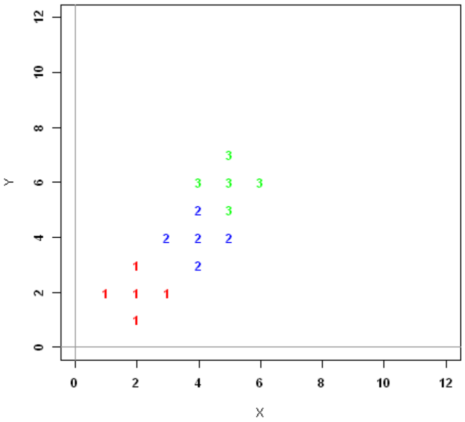
Przykład 4 - $g = 5$

$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ $\bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$



Przykład 4 - $g = 5$

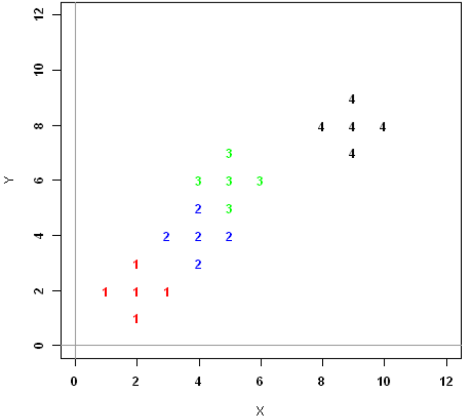
$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$



Przykład 4 - $g = 5$

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

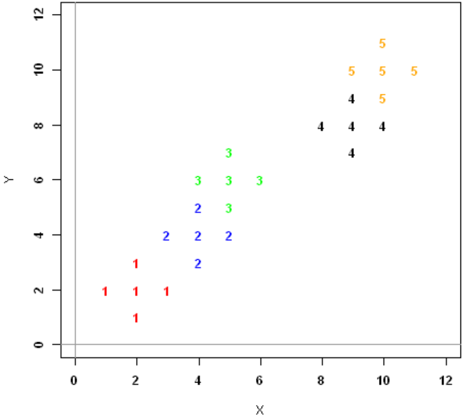
$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$$



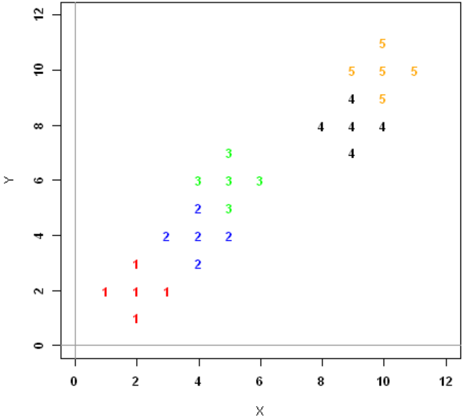
Przykład 4 - $g = 5$

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \quad \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \quad \bar{\mathbf{x}}_5 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$



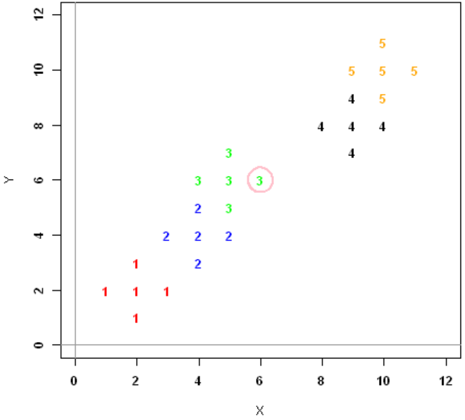
Przykład 4 - $g = 5$



$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \; \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \; \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$
$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \; \bar{\mathbf{x}}_5 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

$$\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{S}_3 = \mathbf{S}_4 = \mathbf{S}_5 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

Przykład 4 - $g = 5$



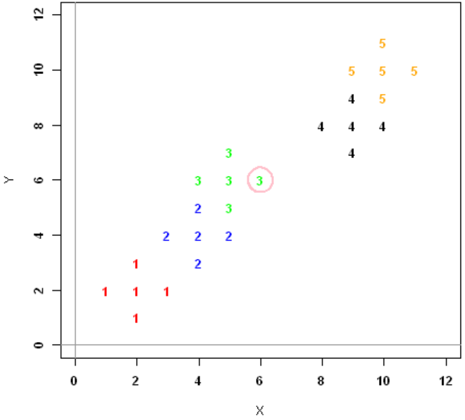
$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \bar{\mathbf{x}}_5 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

$$\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{S}_3 = \mathbf{S}_4 = \mathbf{S}_5 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\bar{\mathbf{x}} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \mathbf{B} = \begin{bmatrix} \frac{115}{2} & \frac{105}{2} \\ \frac{105}{2} & 50 \end{bmatrix}$$

Przykład 4 - $g = 5$



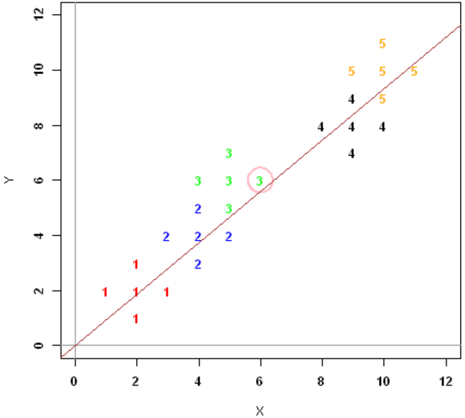
$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$
$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \bar{\mathbf{x}}_5 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

$$\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{S}_3 = \mathbf{S}_4 = \mathbf{S}_5 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\bar{\mathbf{x}} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \mathbf{B} = \begin{bmatrix} \frac{115}{2} & \frac{105}{2} \\ \frac{105}{2} & 50 \end{bmatrix}$$

$$\lambda_1 \approx \frac{425}{2} \; \mathbf{p}_1 = \begin{bmatrix} \frac{15}{14} \\ 1 \end{bmatrix}$$
$$\lambda_2 \approx \frac{5}{2} \; \mathbf{p}_2 = \begin{bmatrix} -\frac{13}{14} \\ 1 \end{bmatrix}$$

Przykład 4 - $g = 5$



$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$
$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \bar{\mathbf{x}}_5 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

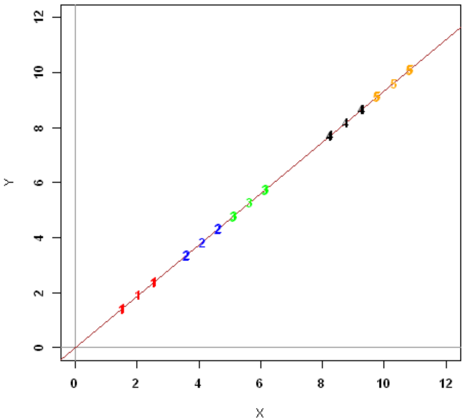
$$\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{S}_3 = \mathbf{S}_4 = \mathbf{S}_5 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\bar{\mathbf{x}} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \mathbf{B} = \begin{bmatrix} \frac{115}{2} & \frac{105}{2} \\ \frac{105}{2} & 50 \end{bmatrix}$$

$$\lambda_1 \approx \frac{425}{2} \ \mathbf{p}_1 = \begin{bmatrix} \frac{15}{14} \\ 1 \end{bmatrix}$$
$$\lambda_2 \approx \frac{5}{2} \ \mathbf{p}_2 = \begin{bmatrix} -\frac{13}{14} \\ 1 \end{bmatrix}$$

$$y = \frac{14}{15} x$$

Przykład 4 - $g = 5$



$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \bar{\mathbf{x}}_5 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

$$\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{S}_3 = \mathbf{S}_4 = \mathbf{S}_5 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

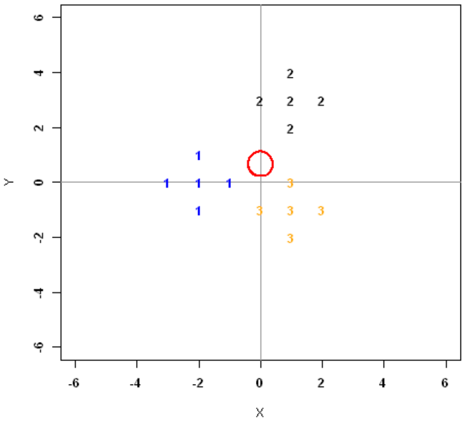
$$\bar{\mathbf{x}} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \mathbf{B} = \begin{bmatrix} \frac{115}{2} & \frac{105}{2} \\ \frac{105}{2} & 50 \end{bmatrix}$$

$$\lambda_1 \approx \frac{425}{2} \; \mathbf{p}_1 = \begin{bmatrix} \frac{15}{14} \\ 1 \end{bmatrix}$$

$$\lambda_2 \approx \frac{5}{2} \; \mathbf{p}_2 = \begin{bmatrix} -\frac{13}{14} \\ 1 \end{bmatrix}$$

$$y = \frac{14}{15} x$$

Przykład 5 - $g = 3$



Przykład 5 - $g = 3$