

Inżynieria Mechatroniczna

Podstawy Sztucznej Inteligencji i Uczenia Głębokiego:
2: Regresja i Klasyfikacja

Ziemowit Dworakowski

AGH w Krakowie

Przykład: Zakup używanego samochodu



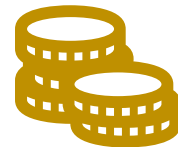
Mamy **7-letniego Opla Astrę**, z **80 000 km przebiegu**. Chcemy sprzedać go **szybko** ale za **jak najwyższą cenę**

Cena początkowa



**Niewiele
zarobimy...**

Za niska

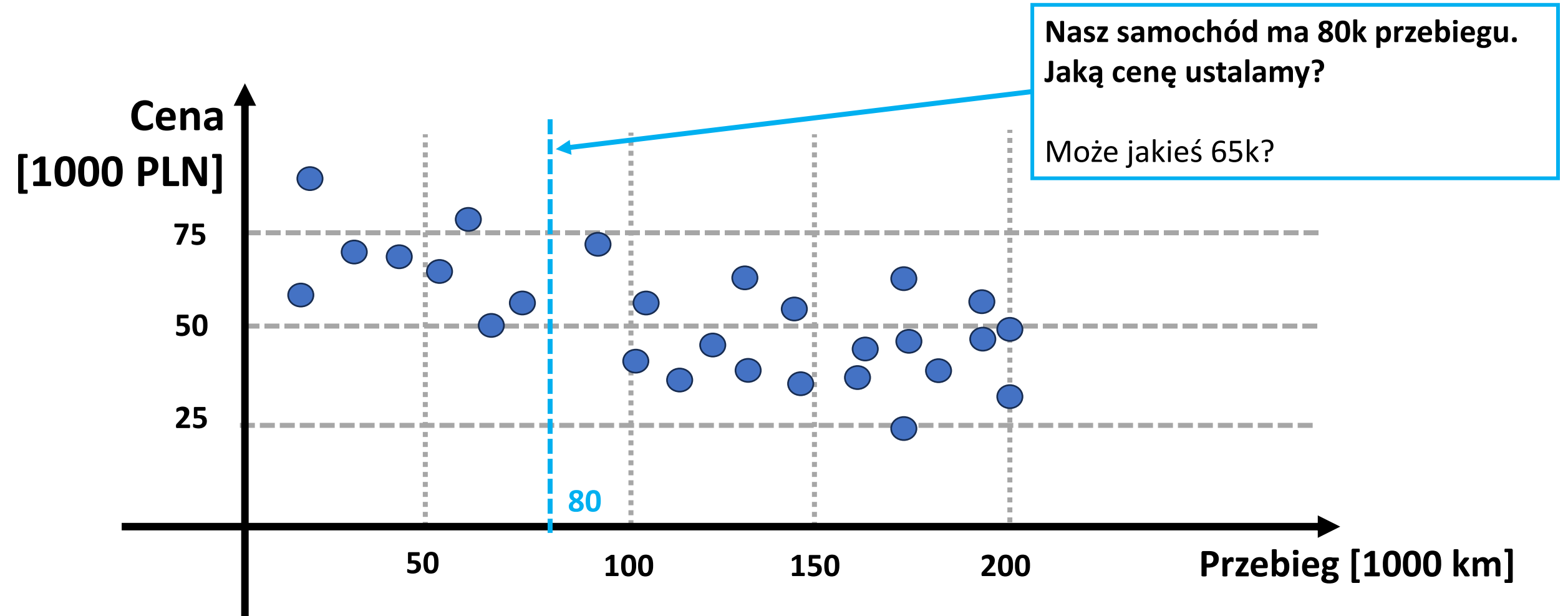


Za wysoka

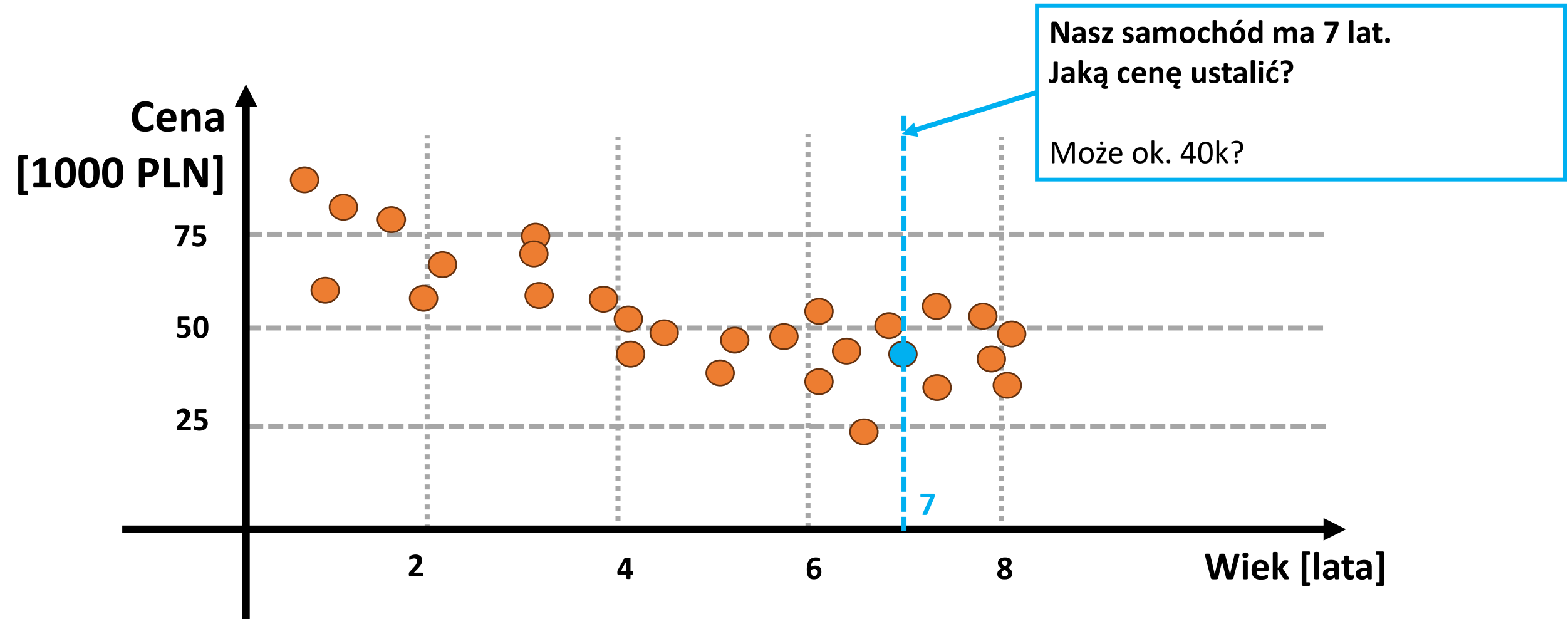


**Samochód nie
sprzeda się szybko**

Przykład: Zakup używanego samochodu

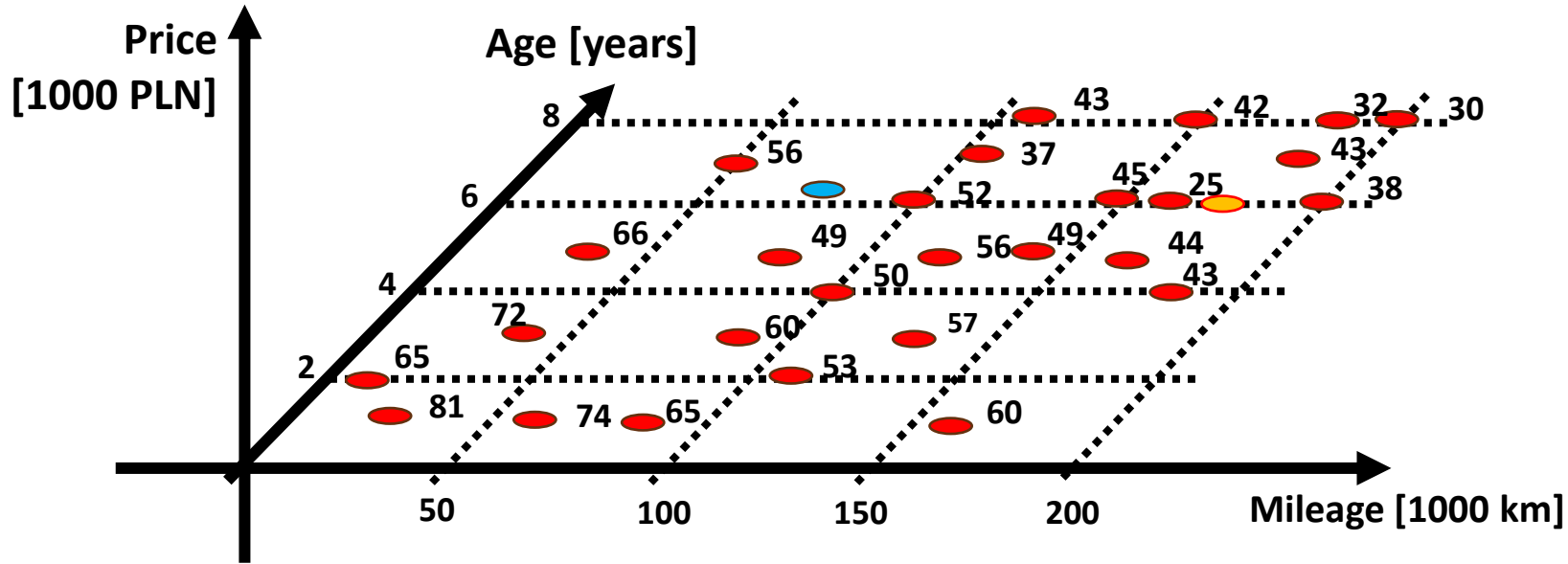


Przykład: Zakup używanego samochodu



Example 1: Used car retail (2023)

ED



Oczywiście nie widzimy wszystkich czynników...

- Samochody mają różne wersje wyposażenia (np. manualna lub automatyczna skrzynia biegów)
- Mogą mieć różny stan
- Na cenę mogą mieć wpływ czynniki geograficzne

Możemy uwzględnić te fakty na kolejnych osiach...



...ale wtedy będziemy potrzebowali więcej danych by gęsto wypełnić przestrzeń...

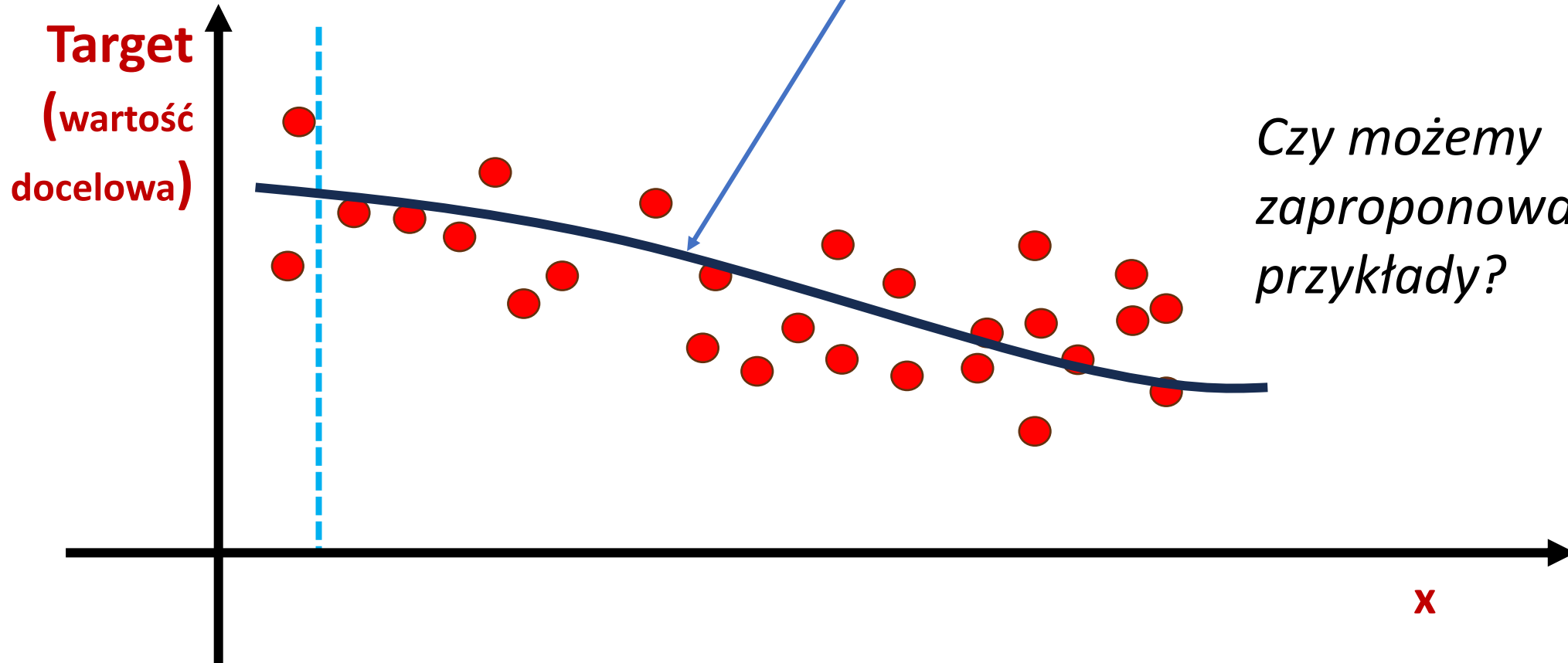
Podsumujmy:

Chcemy zbudować model, który przewiduje **Target** na podstawie **cech**.



$$f(x)$$

Wektor cech
(wejść)



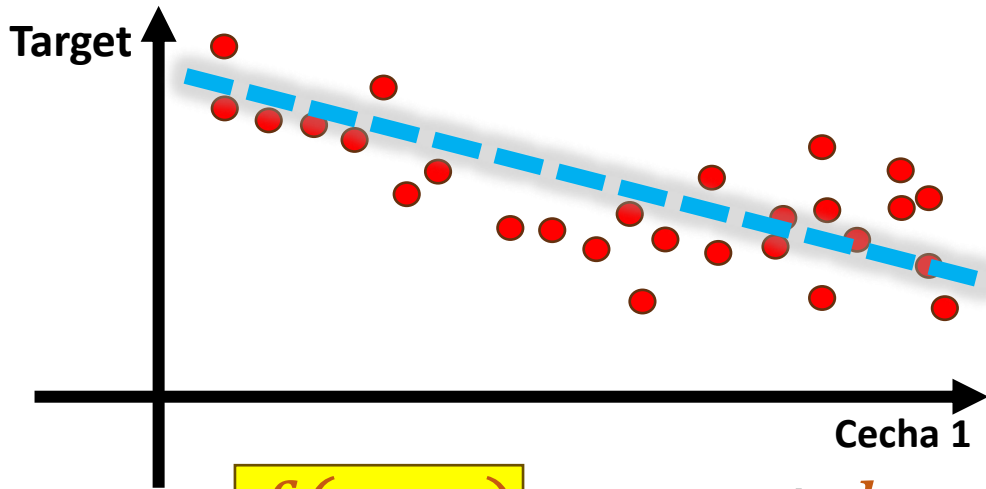
*Czy możemy
zapropionować jakieś inne
przykłady?*

Chcemy zbudować model, który przewiduje **Target** na podstawie **cech i parametrów modelu**.

$$f(w, x)$$

Model liniowy

3D+: używamy **hiperpłaszczyzny**



$$f(w, x) = wx + b$$

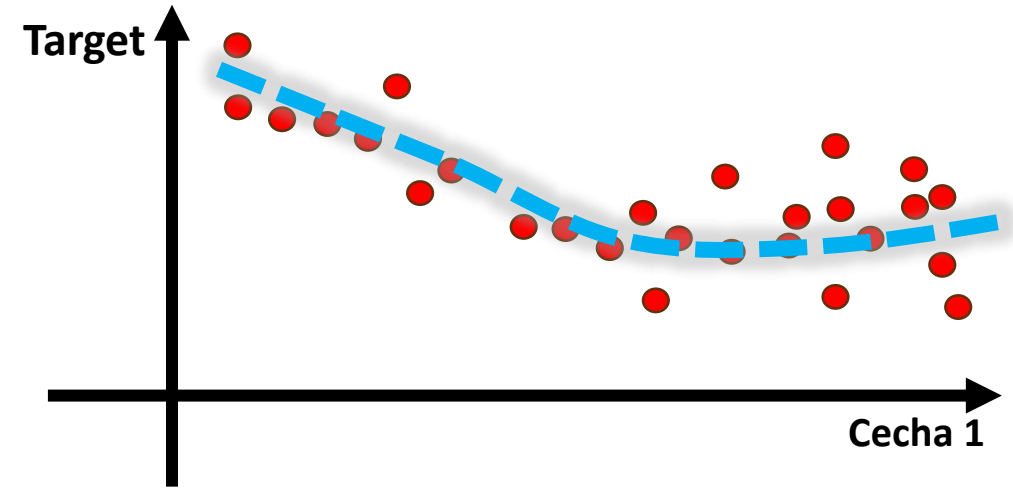
$$w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b$$

■ **Cechy**

■ **Parametry modelu**

Model nieliniowy

3D+: używamy **powierzchni**



Nadal utrzymamy ogólną formę $f(w, x)$, ale teraz zależności będą nieliniowe i zależne od parametrów modelu.

Regresja oznacza poszukiwanie **modelu** estymującego zależność pomiędzy jedną zmienną a pozostałymi

zD.

$$y = f(w, x)$$

Target

inaczej:

Zmienna zależna

Wyjście modelu

Odpowiedź

(Etykieta – w klasyfikacji)

Parametry modelu

inaczej:

Wagi

Nieznane parametry

(Często oznaczane jako β)

Cechy

inaczej:

Zmienne niezależne

Wejścia modelu

Predyktory

Regresja oznacza poszukiwanie **modelu** estymującego zależność pomiędzy jedną zmienną a pozostałymi

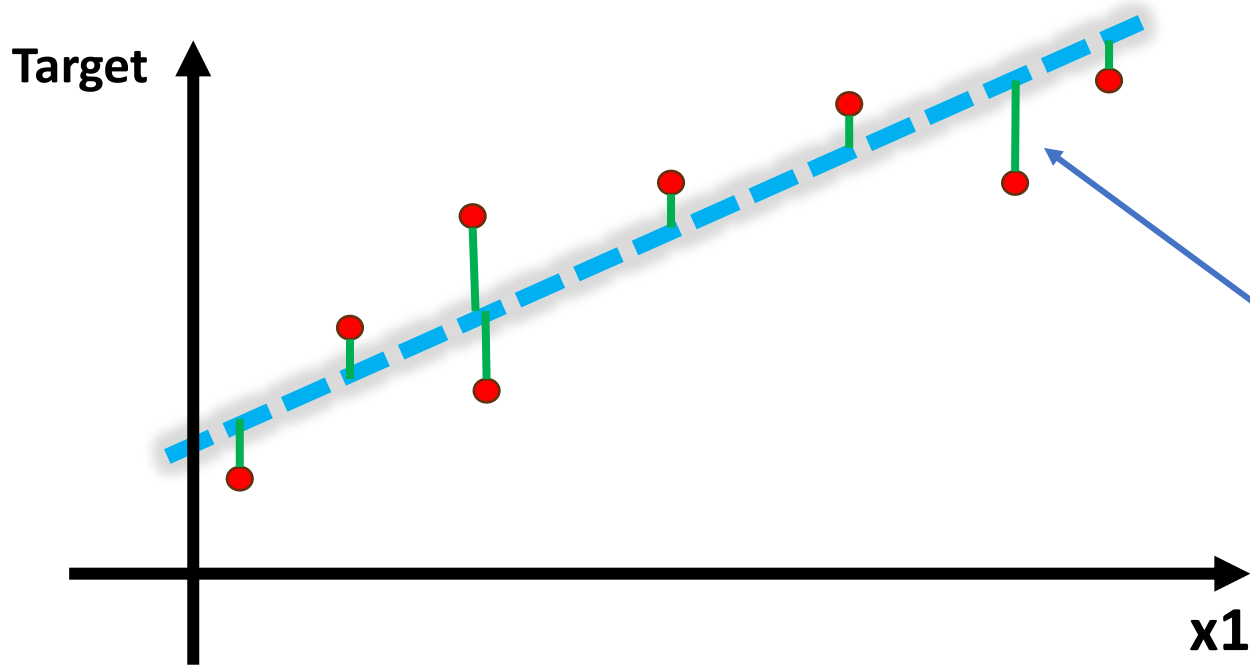
$$y = f(\mathbf{w}, \mathbf{x})$$

Chcemy zminimalizować błąd pomiędzy znaną zmienną zależną Y i wyjściem modelu dla znanego zbioru danych X zawierającego wektory danych \mathbf{x} poprzez dostrajanie parametrów modelu \mathbf{w}

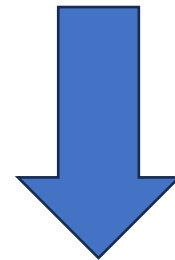
W tym celu możemy użyć błędu średniokwadratowego:

$$\arg \min_{\mathbf{w}} \sum_i (y_i - f(\mathbf{w}, \mathbf{x}_i))^2$$

Regresja liniowa



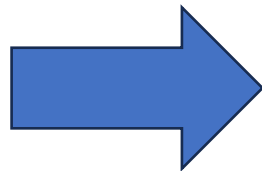
Chcemy znaleźć taką linię, że suma kwadratów zielonych odcinków jest jak najmniejsza



Dopasowanie modelu:

Model:

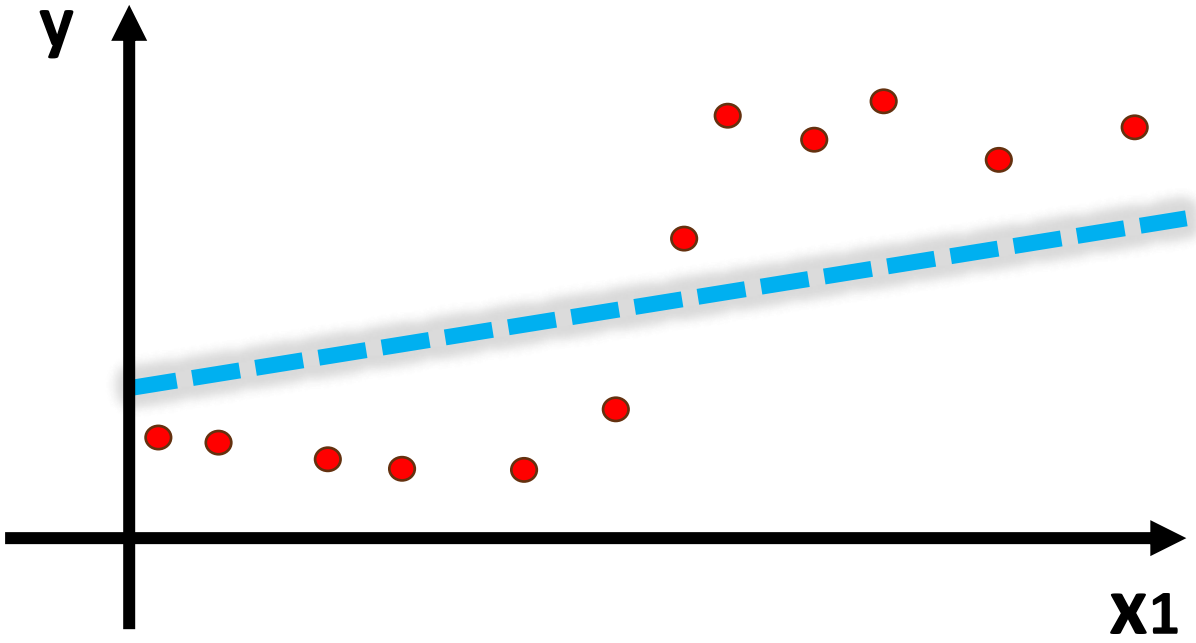
$$y = w_1 x + w_b$$



$$\arg \min_w \sum_i (y_i - (w_1 x_i + w_b))^2$$

Estymator lokalnie liniowy

2D



Linia prosta często nie pozwala na dobre modelowanie danych ...



Ale to na czym nam naprawdę zależy, to skuteczność przewidywania wartości dla konkretnego x

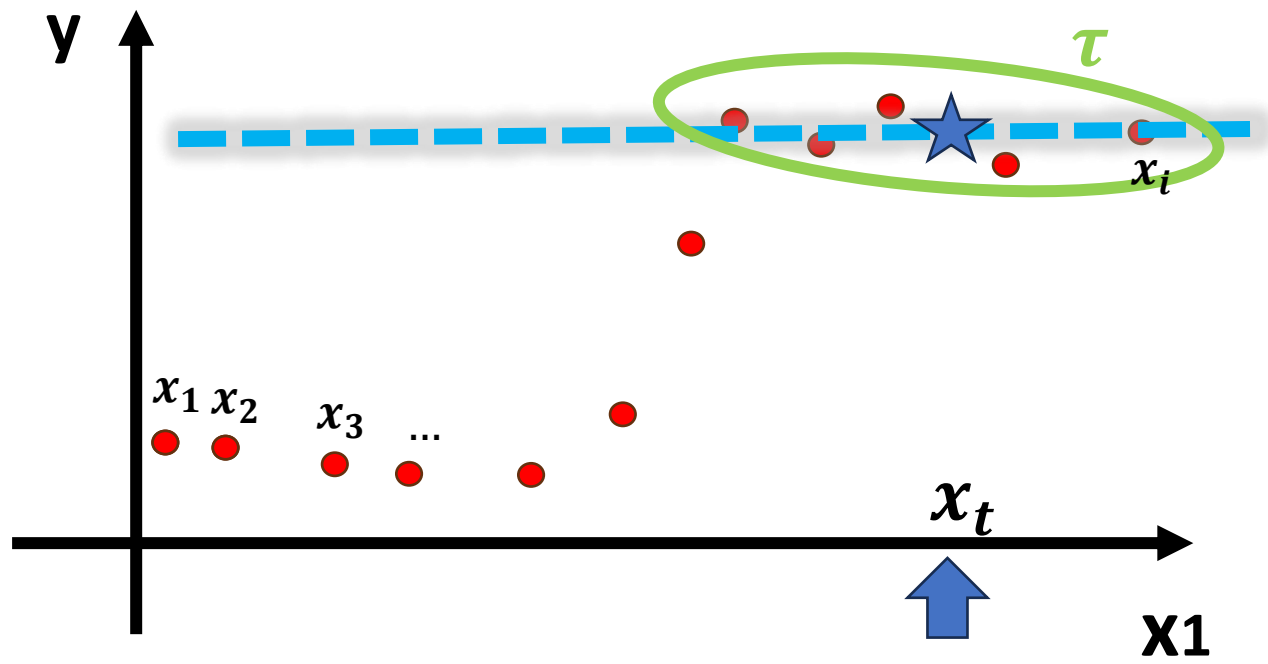


Więc może jesteśmy w stanie zachować model ale zmienić metodę dopasowania?

Estymator lokalnie liniowy

Chcemy zastosować ten sam model: $y = w_1x + w_b$

Tym razem dostroimy go osobno dla każdego przewidywanego punktu wykorzystując informację o sąsiedztwie

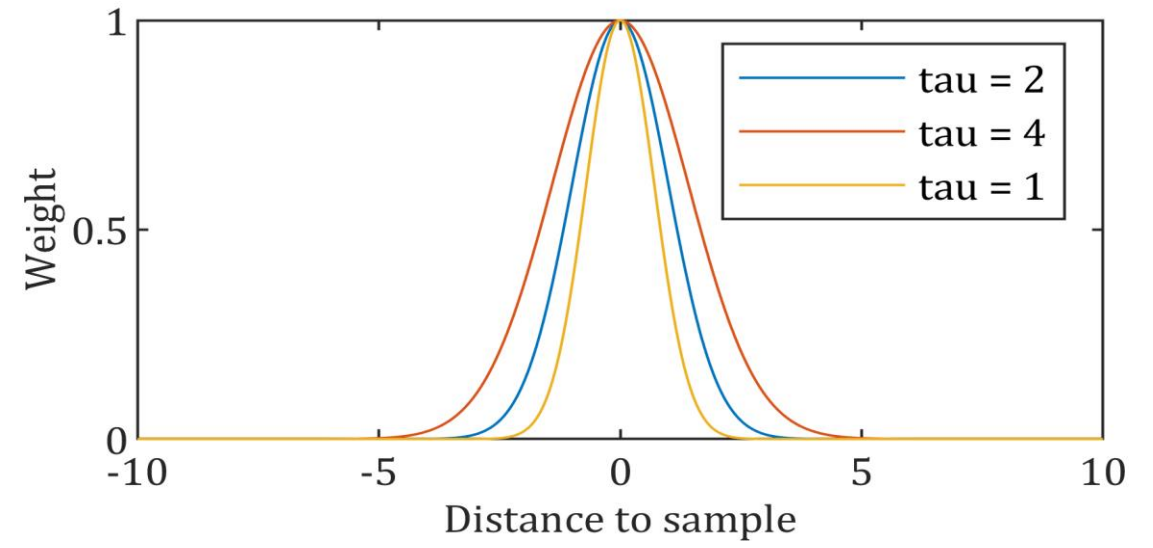


Zastanówmy się nad spodziewanym wyjściem (y) dla tej wartości wejścia (x)

2D

Dla każdego przewidywanego x_t przypisz wagę α_i dla każdego punktu danych x_i gdzie τ służy do określenia wielkości sąsiedztwa

$$\alpha_i = e^{-\frac{(x_i - x_t)^2}{\tau}}$$



$$\arg \min_w \sum_i \alpha_i \cdot (y_i - (w_1x_i + w_b))^2$$

Estymator lokalnie liniowy

2D

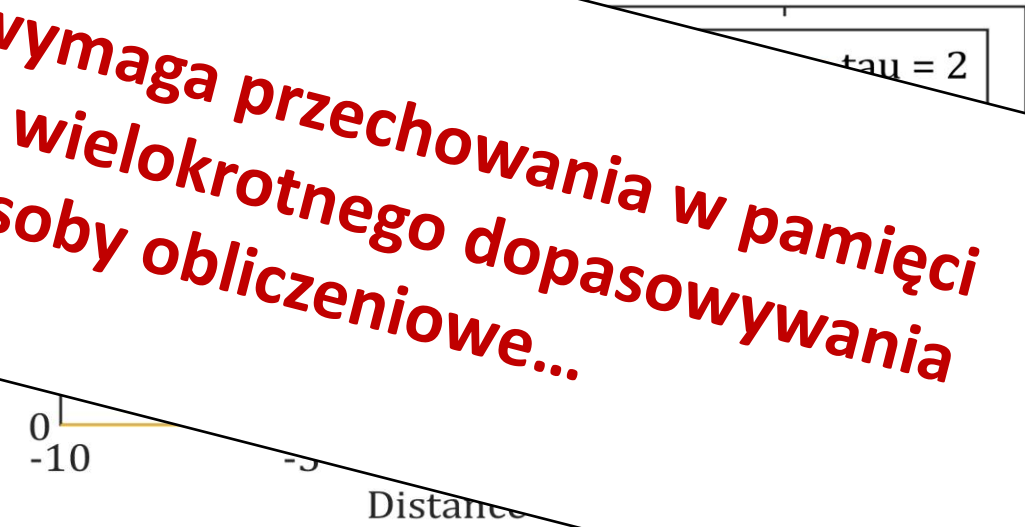
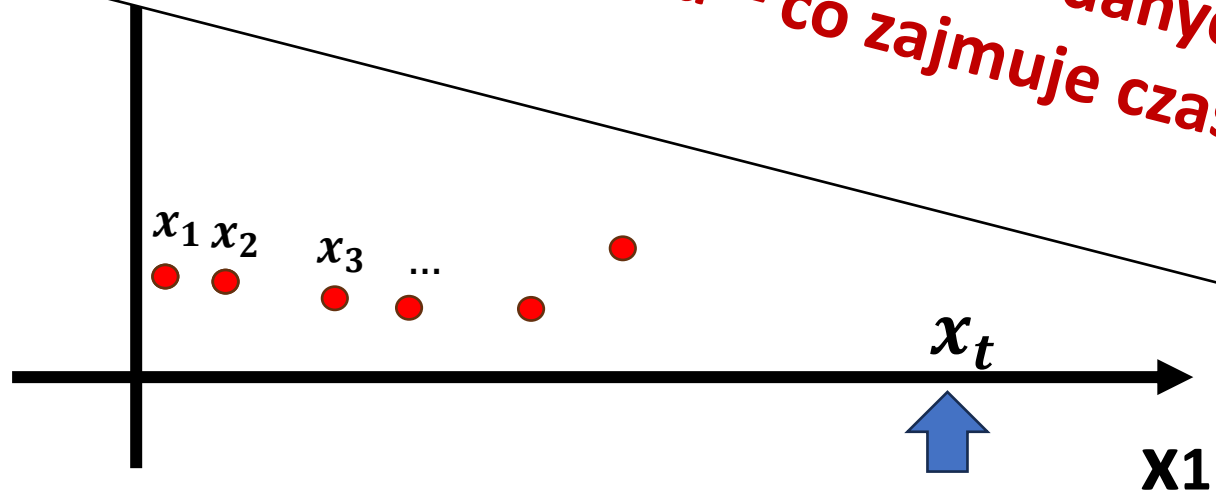
Chcemy zastosować ten sam model: $y = w_1x + w_b$

Tymczasem musimy to zrobić osobno dla każdego punktu danych

Dla każdego przewidywanego x_t przypisz wagę α_i dla każdego punktu danych x_i gdzie τ służy do określenia wielkości sąsiedztwa

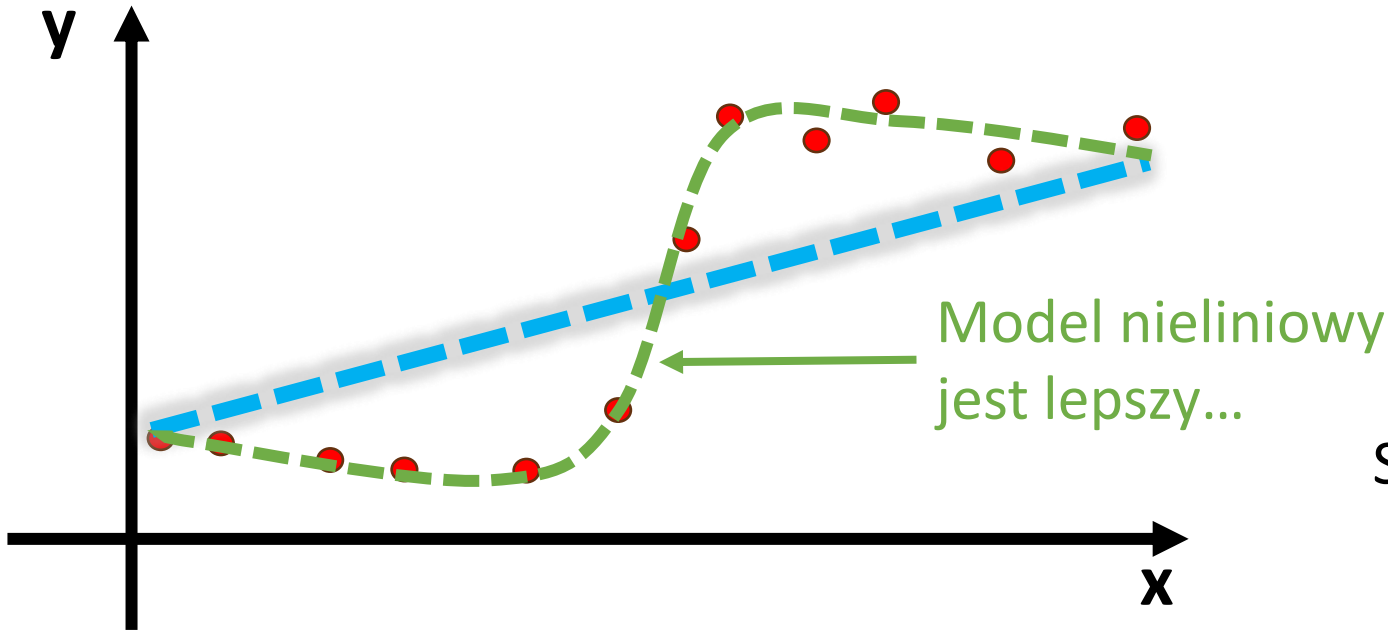
$$\alpha_i = e^{-\frac{(x_i - x_t)^2}{\tau}}$$

Metoda działa bardzo skutecznie – ale wymaga przechowania w pamięci wszystkich punktów ze zbioru danych – i wielokrotnego dopasowywania modelu – co zajmuje czas i zasoby obliczeniowe...



Zastanówmy się nad spodziewanym wyjściem (y) dla tej wartości wejścia (x)

$$\arg \min_w \sum_i \alpha_i \cdot (y_i - (w_1x_i + w_b))^2$$

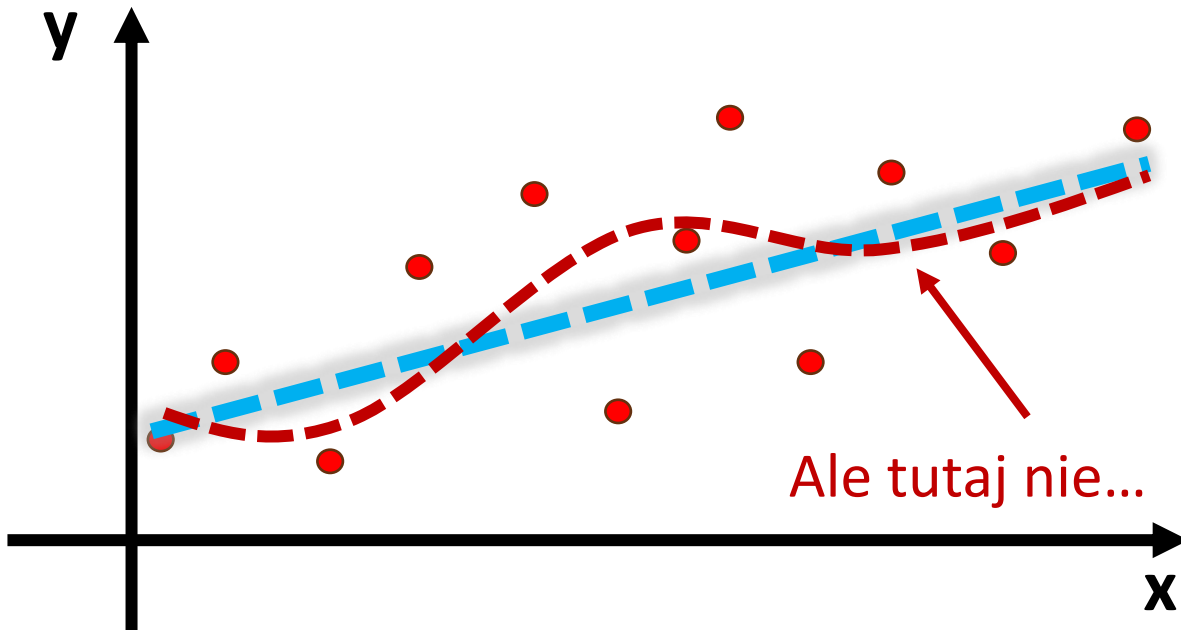


Skąd wiemy, że nieliniowego modelu?

~~Jeśli klasyfikator liniowy ma duży błąd~~



Musimy porównać błąd modelu liniowego i nieliniowego



Model wielomianowy

$$f(\mathbf{w}, \mathbf{x}) = b + w_1 x^1 + w_2 x^2 + w_3 x^3 + \dots$$



Model liniowy



Model kwadratowy

To są wektory, nie pojedyncze wartości!

Modele wielomianowe działają dokładnie tak samo jak liniowe. Jedyna różnica to konieczność ustawienia dodatkowego **metaparametru** – stopnia użytego wielomianu

Robimy to zazwyczaj poprzez zwiększanie tego stopnia aż model przestanie się poprawiać...

Czy to wszystko?

Gaussian mixture models

Algorytm komitetowy Regresja Bayesowska

Maszyna Wektorów Podtrzymujących (SVM)

Multivariate Adaptive Regression Splines

Regresja kNN

Sieci konwolucyjne

Bayesowskie sieci neuronowe

Sztuczne Sieci Neuronowe (ANNs) *Transformery*

Wielowarstwowe perceptrony

Sieci Echo State

Sieci radialne

Long Short-Term Memory networks (LSTMs)

Extreme learning machines

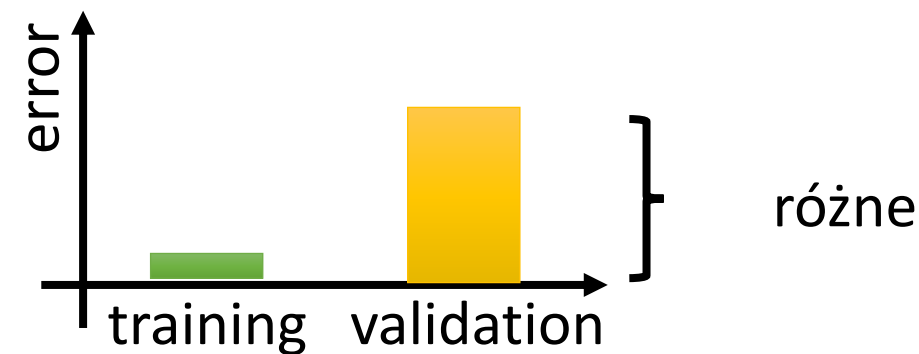
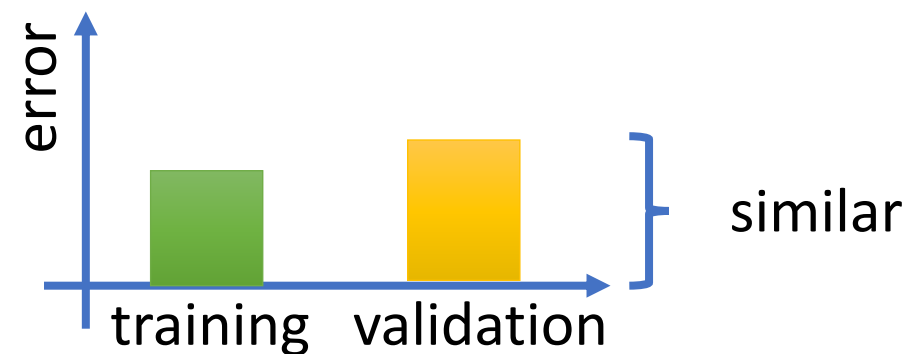
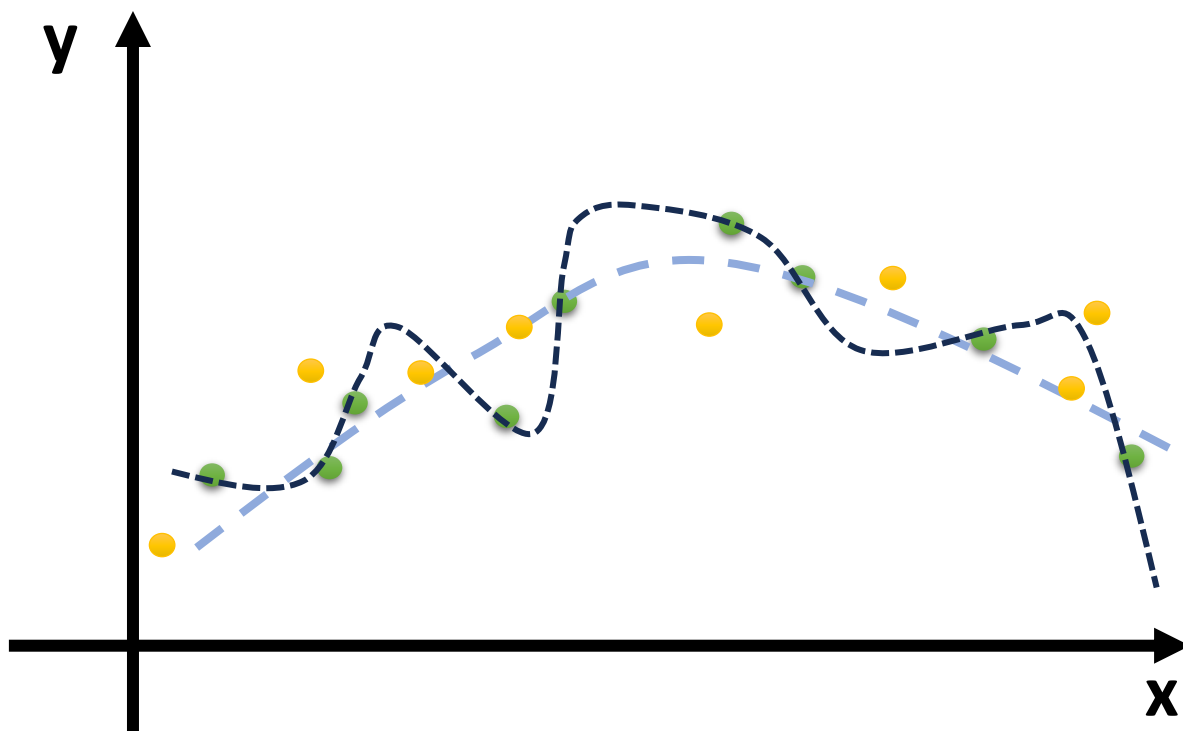
Sieci rekurencyjne

...część z nich poznamy na pozostałej części przedmiotu

Przeuczenie

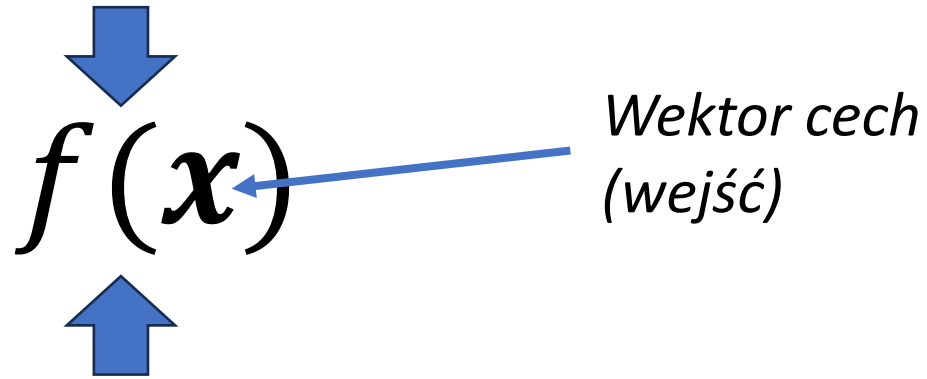
Przeuczenie oznacza, że model zapamiętuje dane treningowe kosztem zdolności generalizacyjnych (zdolności uogólniania)

Rozpoznamy ją poprzez porównanie skuteczności treningowej i skuteczności na niezależnym **walidacyjnym** podzbiore danych. Jeśli znacząco spada – model jest przeuczony.



Regresja:

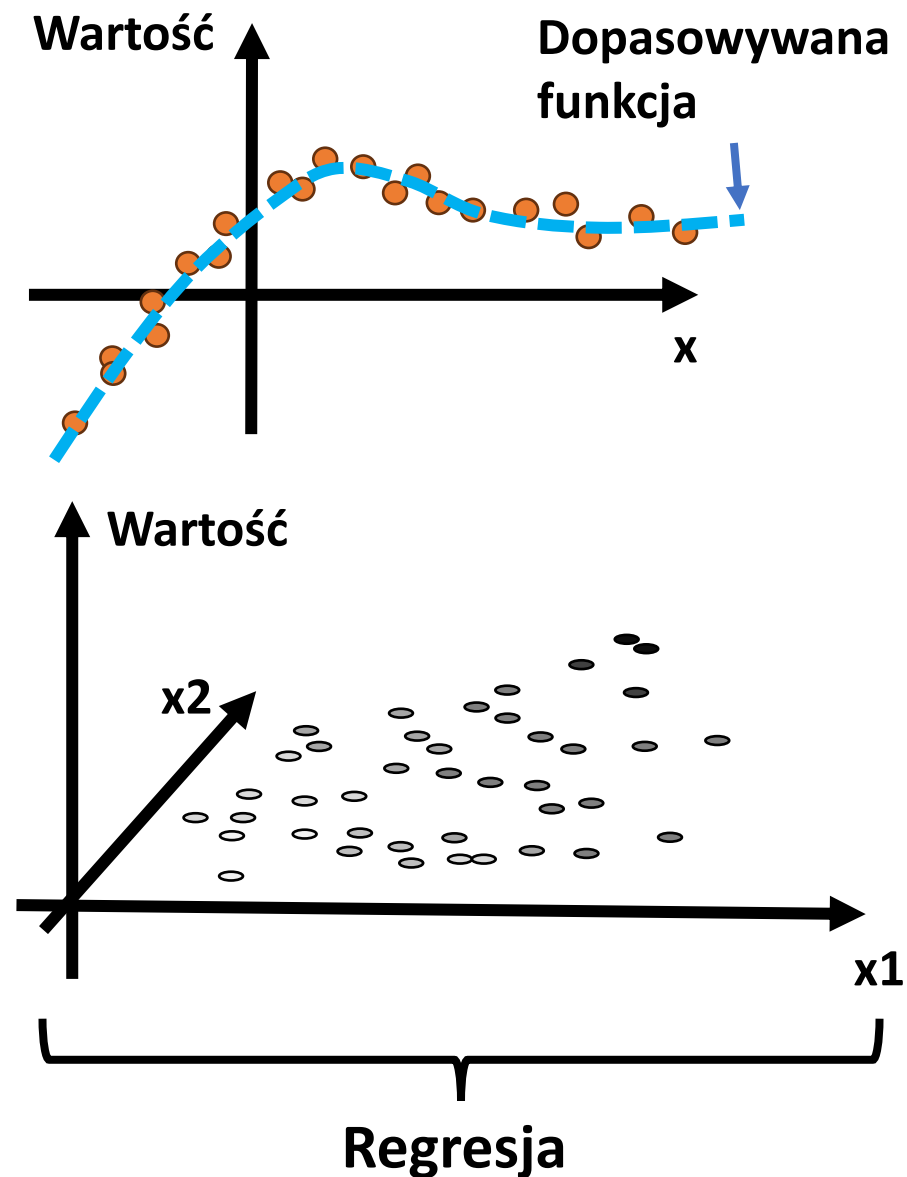
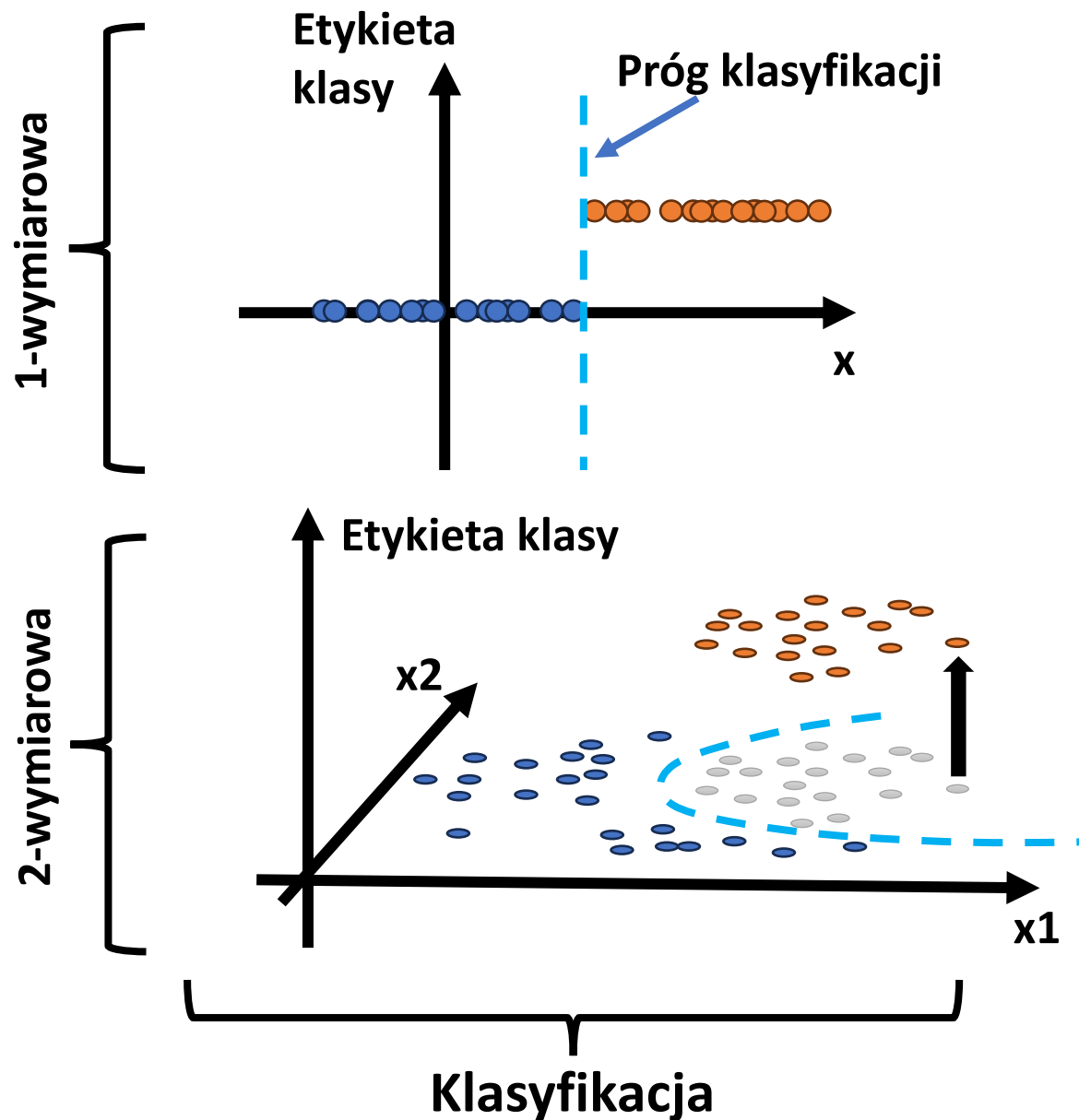
Chcemy zbudować model, który przewiduje **Target** na podstawie **cech**.



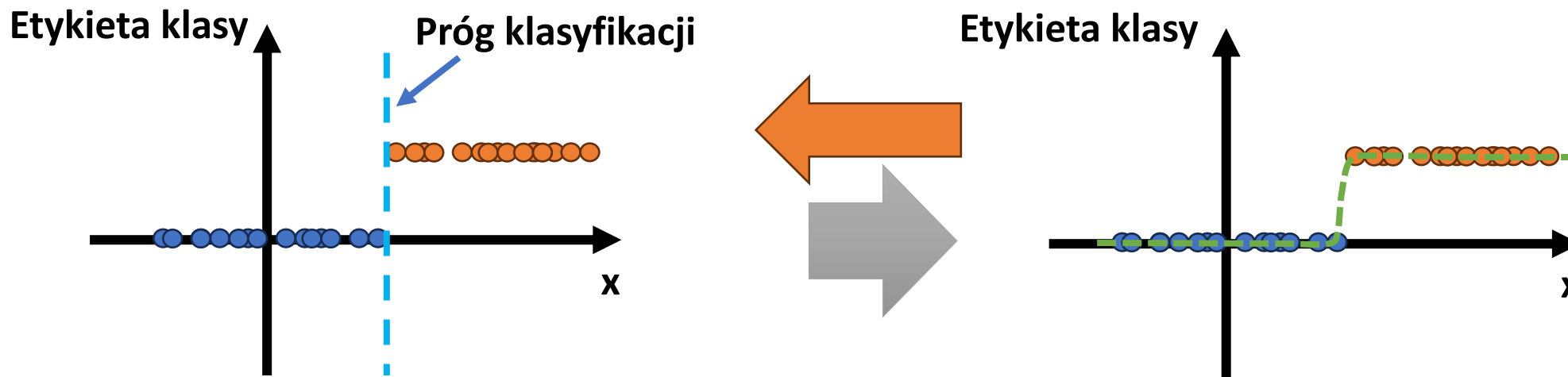
Klasyfikacja:

Chcemy zbudować model, który przewiduje **Etykietę** na podstawie **cech**.

Analogia między klasyfikacją i regresją



Analogia między klasyfikacją i regresją



Możemy zinterpretować klasyfikację jak regresję w której celem jest odnalezienie funkcji kodującej etykietę klasy

W tym celu możemy (tak jak w regresji) minimalizować błąd pomiędzy znanymi etykietami i wyjściem modelu

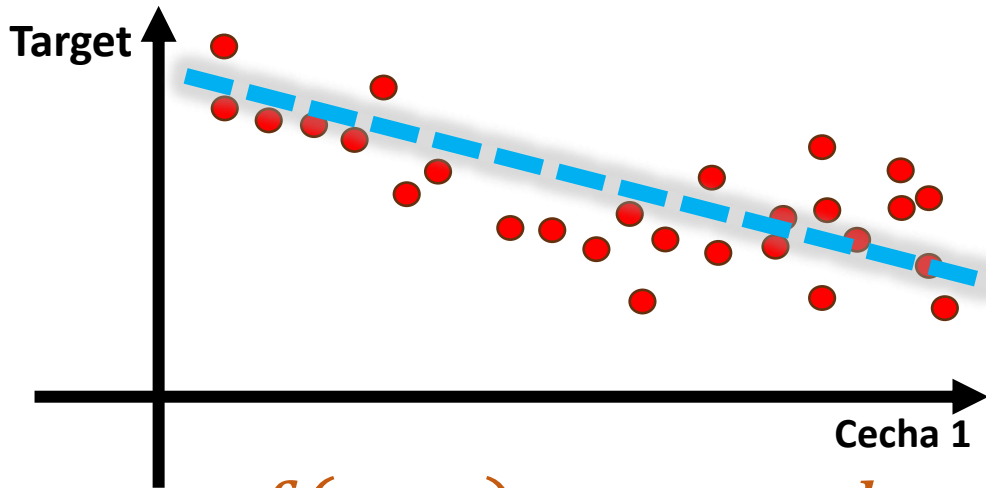
Klasyfikacja:

Chcemy zbudować model, który przewiduje **Etykietę klasy** na podstawie **cech i parametrów modelu**.

3D

Model liniowy

3D+: używamy **hiperpłaszczyzny**



$$f(w, x) \supseteq wx + b$$

$$w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b$$

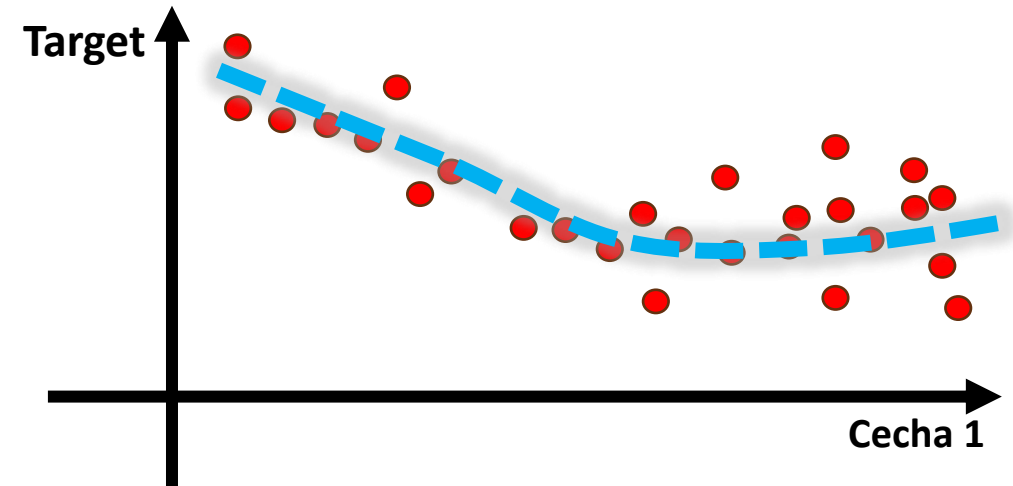
■ Cechy

■ Parametry modelu

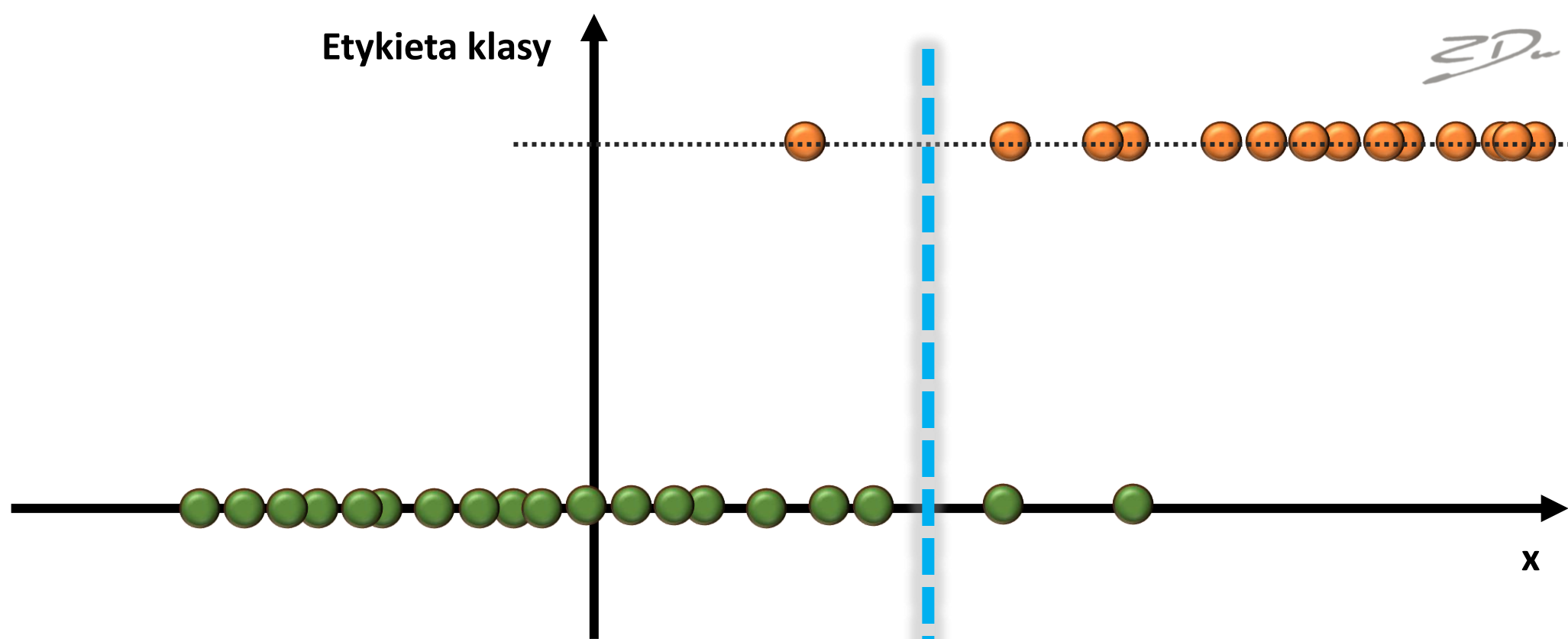
$$f(w, x)$$

Model nieliniowy

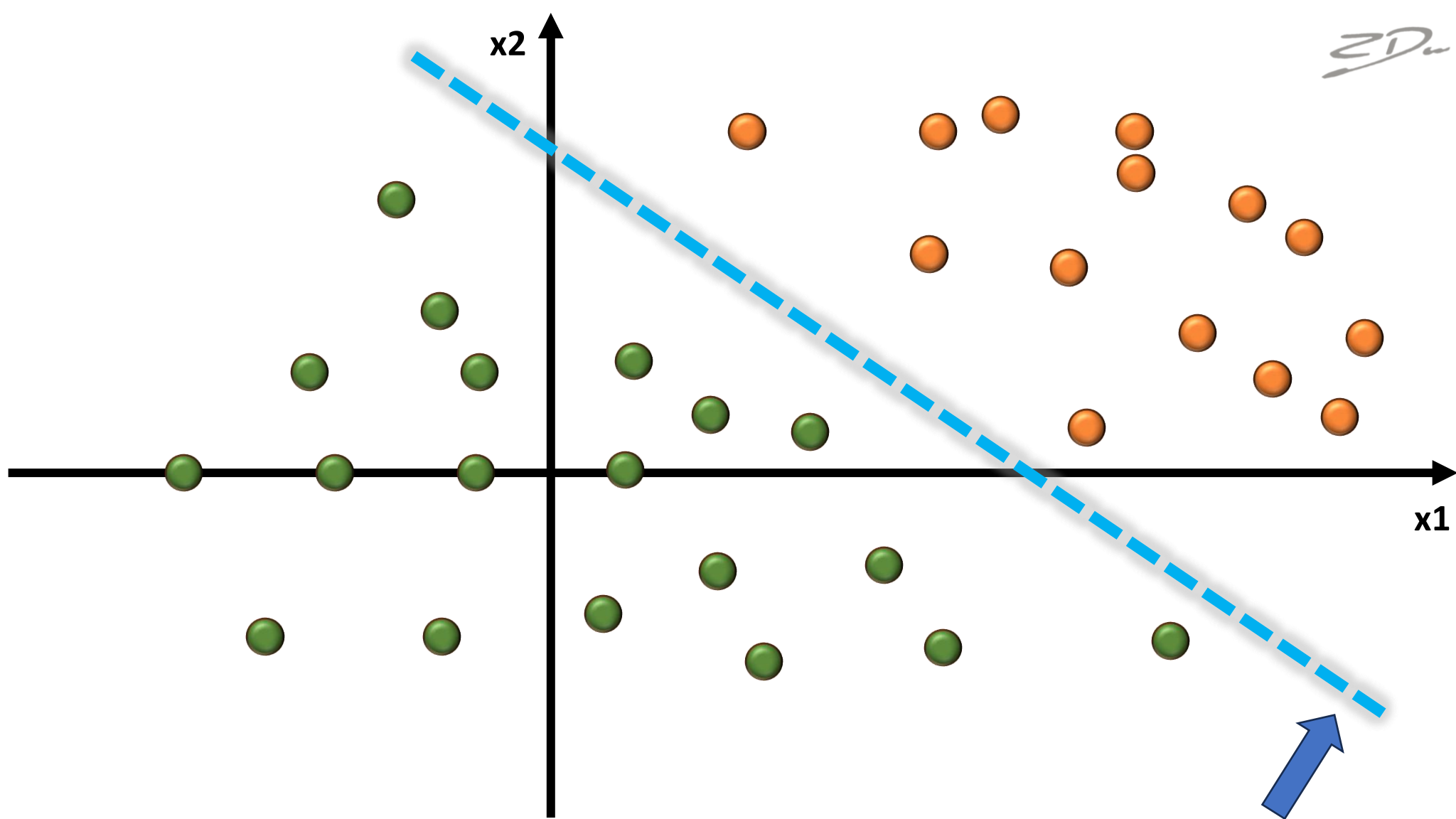
3D+: używamy **powierzchni**



Nadal utrzymamy ogólną formę $f(w, x)$, ale teraz zależności będą nieliniowe i zależne od parametrów modelu.



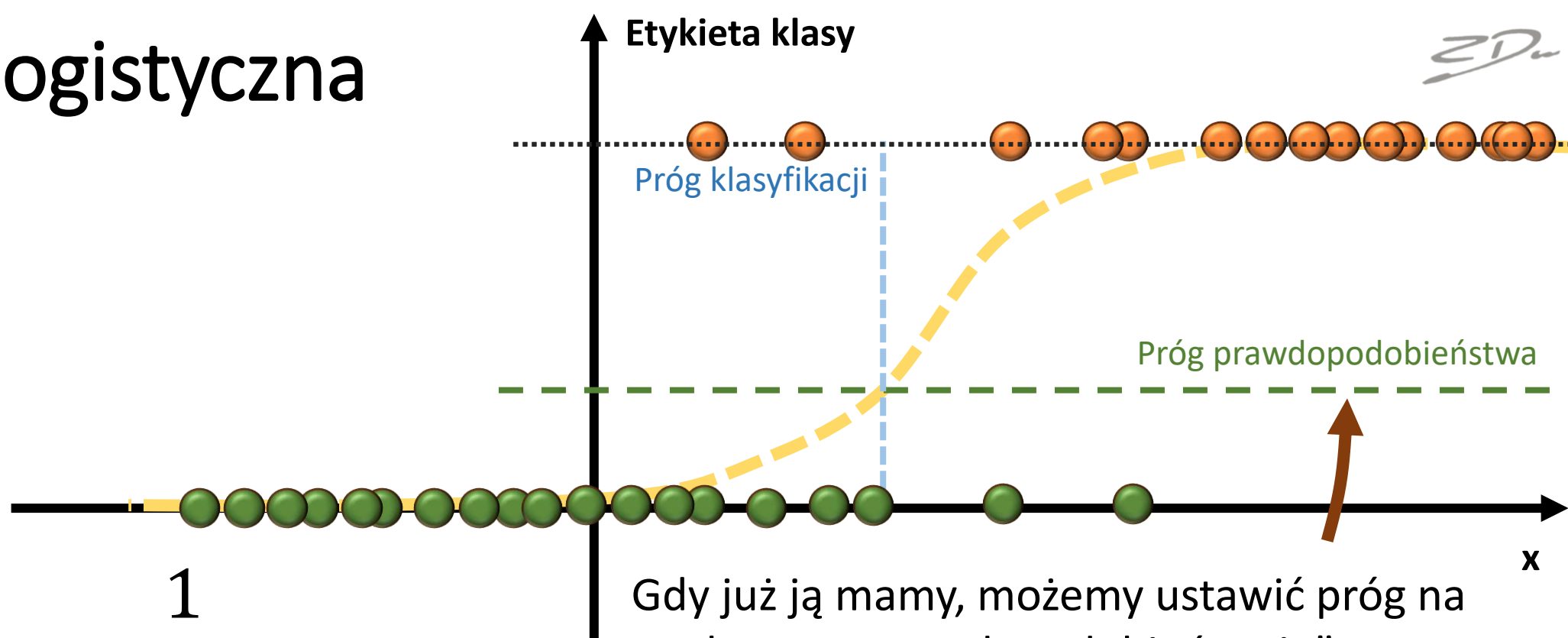
Modyfikujemy położenie progu klasyfikacji
by zminimalizować **liczbę błędów** (lub
prawdopodobieństwo błędu)



Modyfikujemy położenie progu klasyfikacji
by zminimalizować **liczbę błędów** (lub
prawdopodobieństwo błędu)

Regresja logistyczna

2D



$$f(x) = \frac{1}{1 + e^{-(\beta_1 x + \beta_0)}}$$

Gdy już ją mamy, możemy ustawić próg na „wybranym prawdopodobieństwie”

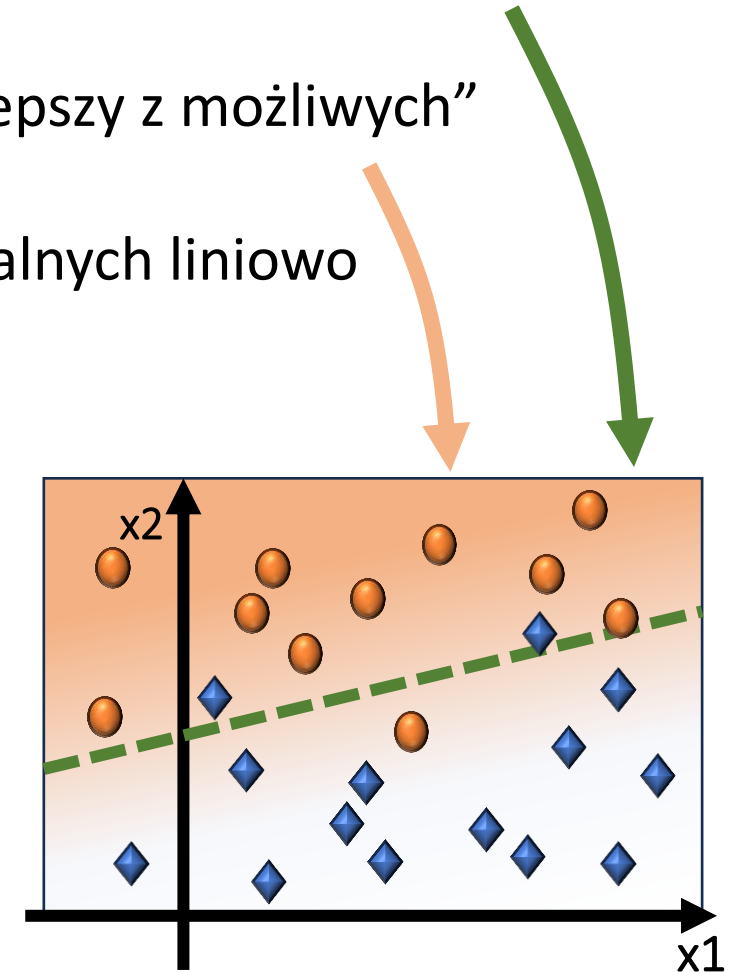
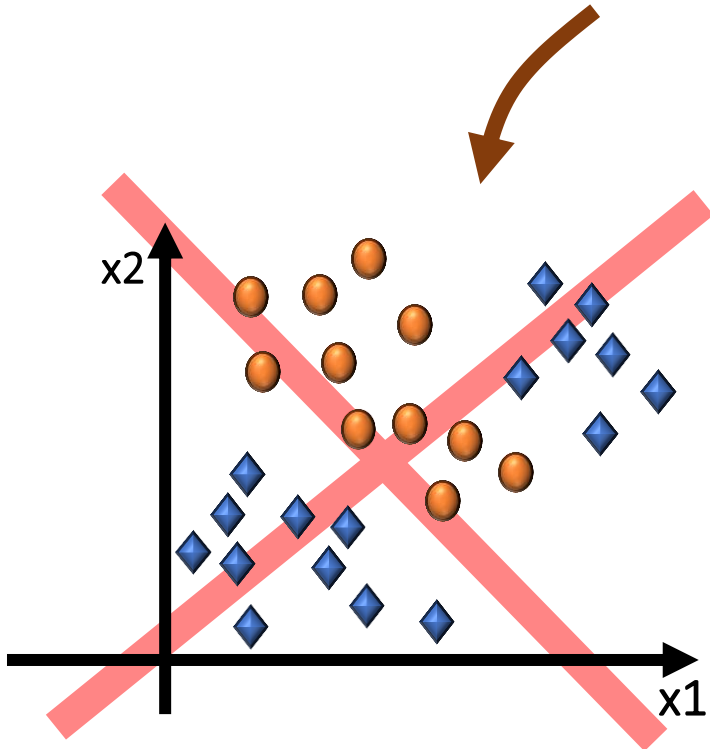
To krzywa „sigmoidalna” – tutaj nazywana „funkcją logistyczną”.
 β_1 i β_0 to parametry które muszą być dopasowane tak, by zmaksymalizować logarytm prawdopodobieństwa poprawnego przypisania etykiety.

Możemy w tym celu zastosować np. algorytm gradientowy

Regresja logistyczna

SD

- LR to klasyfikacyjny odpowiednik regresji liniowej
- Dobrze działa również w przestrzeni wielowymiarowej (mamy jedynie więcej parametrów do nauczenia)
- Może wpasować „klasyfikator liniowy” w sposób „najlepszy z możliwych” (tzn. klasyfikacja ma interpretację probabilistyczną)
- Nie jest w stanie rozwiązywać problemów nieseparowalnych liniowo



Materiał do powtórki:



1. Wyjaśnij na przykładzie zadanie regresji (w tym wielowymiarowej)
2. Wyjaśnij algorytmy regresji liniowej, wielomianowej, estymatora lokalnie liniowego. Podaj ich właściwości (wady, zalety)
3. Wyjaśnij czym jest przeuczenie i jak można je wykryć
4. Wyjaśnij czym się różni klasyfikacja od regresji (w tym na przykładach)
5. Wyjaśnij zasadę działania regresji logistycznej, podaj jej zalety i wady i podaj wzór na funkcję logistyczną