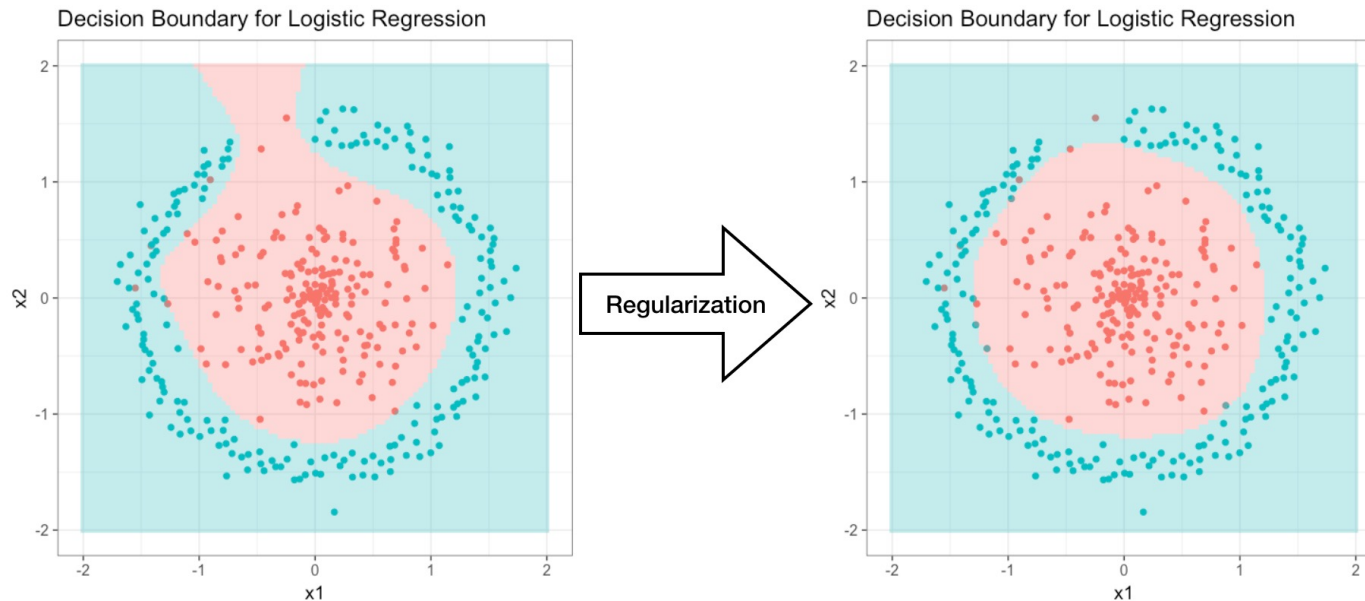


Regularyzacja

- Model przeuczony bardzo kiepsko sprawdzi się na nieobserwowanych dotąd danych.
- Potrzeba mechanizmu, który potrafi zapobiegać przeuczeniu, 'sterować' tym na ile model ma generalizować dane trenujące.
- Do funkcji kosztu (na etapie trenowania modelu) dodajemy komponent regularyzacyjny.



Regularyzacja

- Jak określić $R(\theta)$?
- Jak wybrać odpowiednią wartość dla λ ?

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{n} \sum_{i=1}^n \mathbf{Loss}(y_i, f_{\theta}(x_i)) + \lambda \mathbf{R}(\theta)$$

Fit the Data

Penalize
Complex Models

Regularization
Parameter

Regularyzacja L2 – Ridge Regression

- W przypadku zwykłej regresji liniowej szukane są wagi minimalizujące funkcję celu postaci:

$$J(w) = \sum_{i=1}^m (y_i - w^T x_i)^2$$

- Dla regresji grzbietowej - Ridge regression lub Thikonov - dodana jest funkcja kary, czyli składnik ograniczający wartości współczynników:

$$J(w) = \sum_{i=1}^m (y_i - w^T x_i)^2 + \lambda \|w\|^2$$

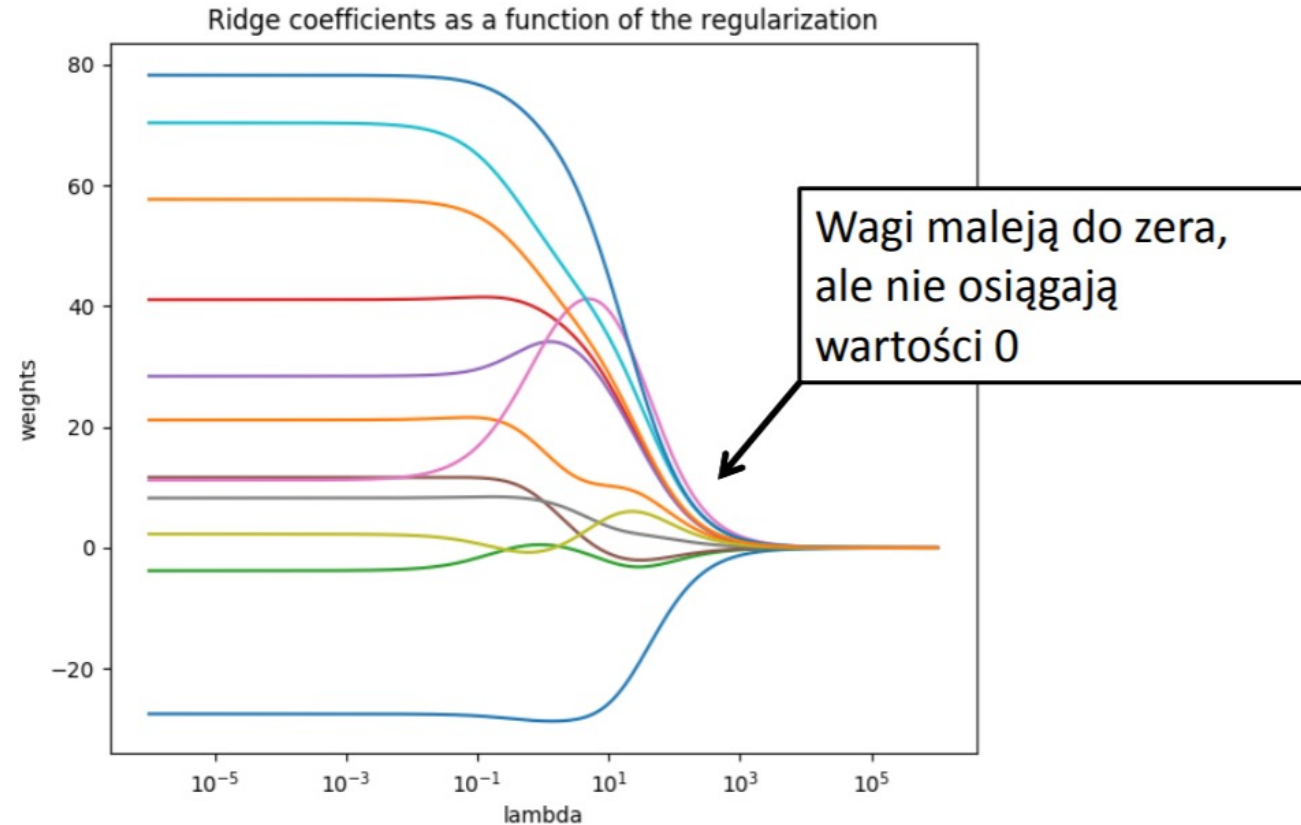
- Czynnik $\|w\|^2$ to tzw. norma L^2 , czyli norma euklidesowa:

$$w^T w = \sum_{i=1}^n w_i^2$$

- Learning rate, czyli λ określa udział kary w funkcji celu (niekiedy learning rate oznaczany jest jako α)

Regularyzacja L2 – Lambda

- Jeżeli $\lambda = 0$, funkcja celu jest taka sama, jak dla zwykłej regresji
- Dla małych wartości λ wpływ czynnika regularyzującego będzie mniejszy – współczynniki będą się powiększać
- Dla dużych wartości λ współczynniki będą bliskie zeru (i większości przypadków błąd RSS będzie duży)



Regularyzacja L1 – Lasso

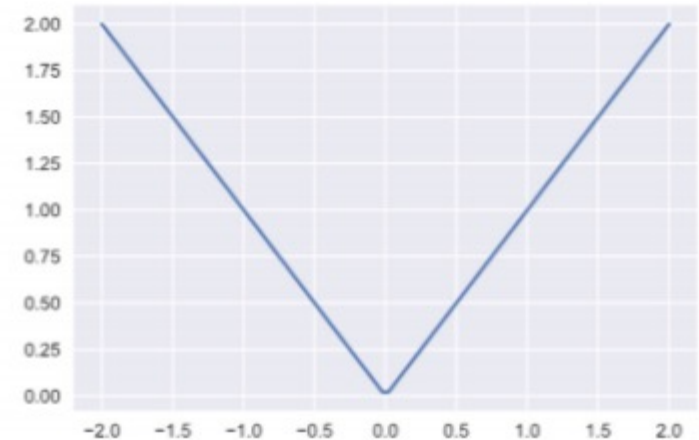
- W przypadku regularyzacji L1 składnikiem regularyzującym jest norma L1 (czyli suma wartości bezwzględnych wag)

$$J(w) = \sum_{i=1}^m (y_i - w^T x_i)^2 + \lambda \sum_{i=1}^n |w_i|$$

- Lasso może sprowadzić wartości wag do 0, przez co ten typ regularyzacji działa jak mechanizm wyboru atrybutów (feature selection) - stopniowo odrzuca współliniowe atrybuty, pozostawia zbiór najbardziej istotnych (tych, które najlepiej „objaśniają” zmienność wartości wyjściowych).

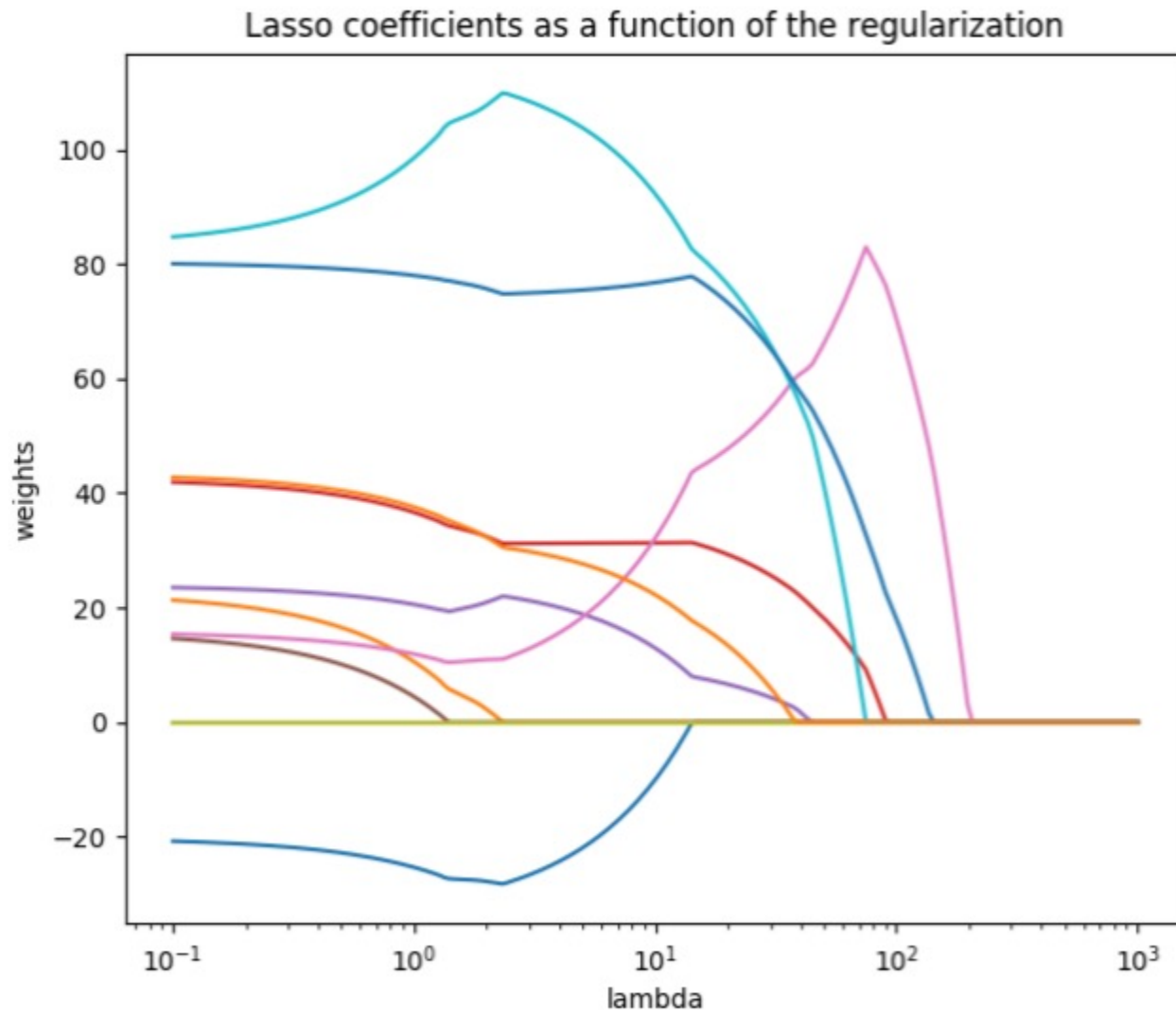
LASSO
(L1-Reg)

$$R_{\text{Lasso}}(\theta) = \sum_{i=1}^d |\theta_i|$$



Regularyzacja L1 – Lambda

- Dla L1 wraz ze wzrostem λ kolejne współczynniki będą znikać (przyjmować wartość 0)
- Dla L2 wagi będą stawały się dowolnie małe, ale nie zanikały zupełnie



Regularyzacja Elastic Net

(Elastic Net regression)

- Łączy regresję grzbietową z Lasso

$$J(w) = \frac{\sum_{i=1}^m (y_i - w^T x_i)^2}{2m} + \lambda \left(\frac{1 - \alpha}{2} \sum_{i=1}^n w_i^2 + \alpha \sum_{i=1}^n |w_i| \right)$$

Lambda – jak wybrać parametr

- Stosując metodę walidacji krzyżowej: dla siatki różnych wartości λ obliczamy błąd walidacji krzyżowej dla tego wyboru λ .
- Wybieramy tę wartość λ , dla której ten błąd jest najmniejszy
- Trenujemy model już na całych danych, z wybranym λ .