

Regresja w ML

Modele regresyjne są używane do predykcji ciągłych wartości. Są to techniki typu supervised.

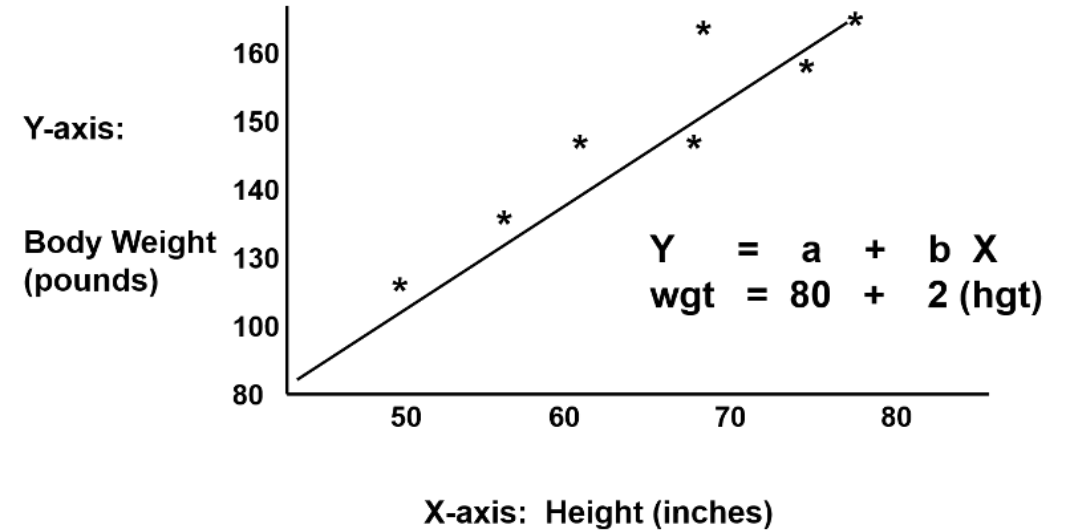
1. Simple Linear Regression
2. Polynomial Regression
3. Support Vector Regression
4. Decision Tree Regression
5. Random Forest Regression

Regresja liniowa

Celem jest znalezienie takich wartości współczynników, które minimalizują cost functions. Najczęściej używaną cost functions jest Mean Squared Error (MSE).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Wartości współczynników równania regresji mogą być wyznaczone za pomocą metody gradient descent.

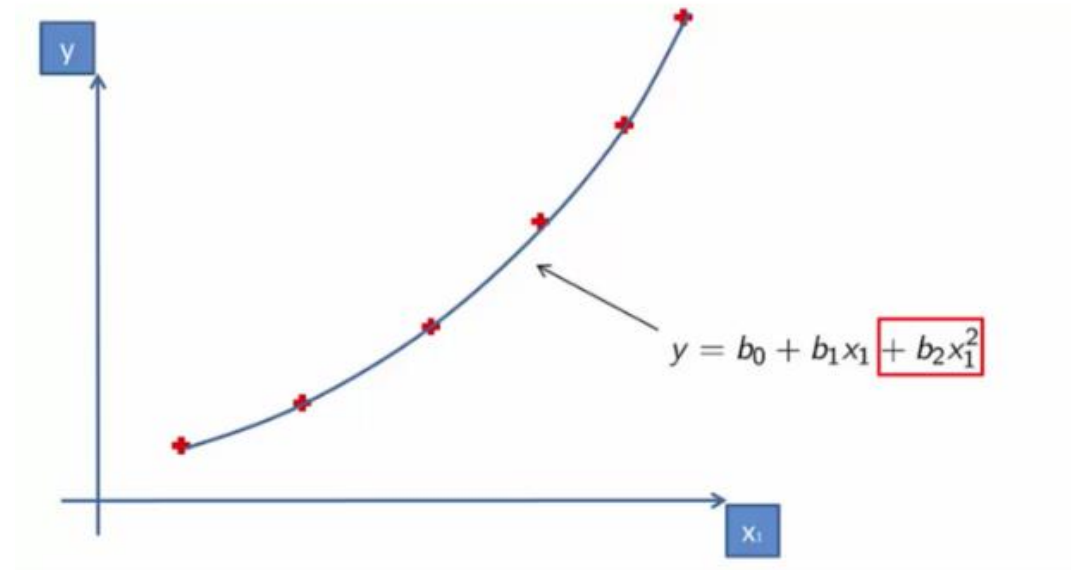


Regresja wielomianowa (polynomial)

W regresji wielomianowej, przeprowadza się transformację oryginalnych features do wielomianowych features danego stopnia i następnie stosuje się regresję liniową.

Model liniowy jest transformowany do postaci:

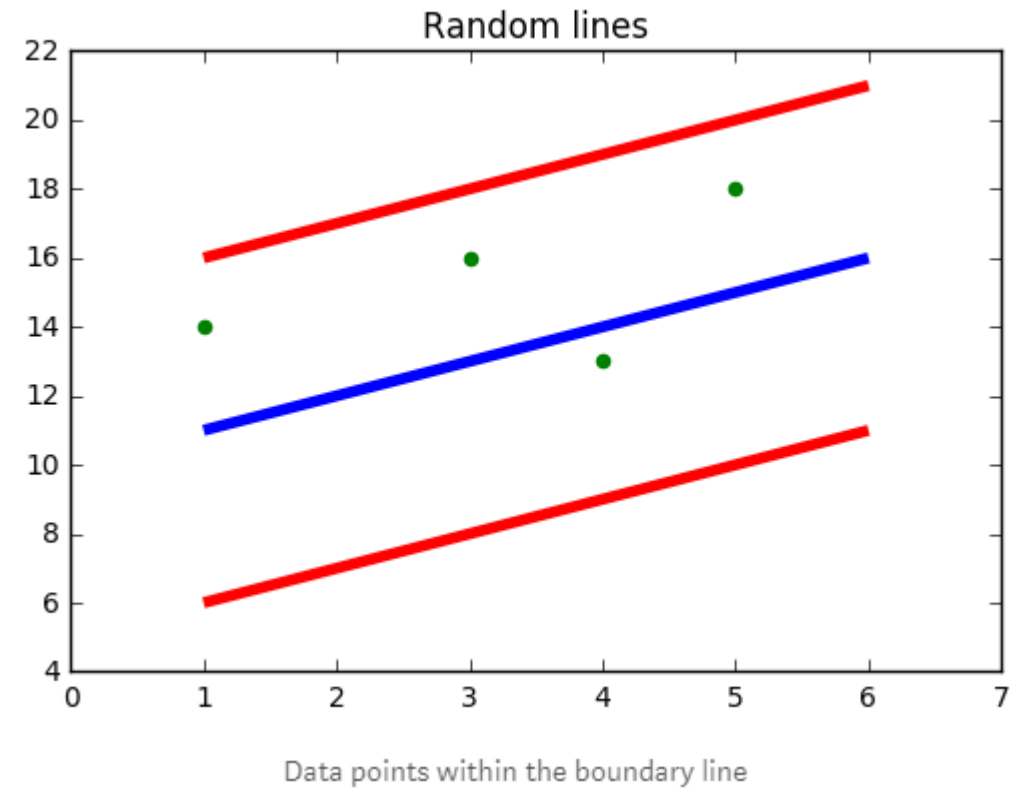
$$Y = a + bX \quad \longrightarrow \quad Y = a + bX + cX^2$$



Zwiększanie stopnia do dużych wartości, prowadzić może do overfitting

Support Vector Regression

W **SVR** identyfikujemy hyperplane , staramy się dopasować ją tak, aby różnice pomiędzy wartościami prawdziwymi i prognozowanymi (errors) nie przekraczały pewnej granicy (w zakresie pewnej threshold). Najlepszym dopasowaniem jest hyperplane o maksymalnej liczbie punktów.



Decision Tree Regression

Są używane zarówno do klasyfikacji jak i regresji. Podstawą jest na każdym poziomie drzewa określenie splitting attribute. Odbywa się to (ID3) przez redukcję odchylenia standardowego.

Czyli dokonuje się podziału wartości na podobne (jak najbardziej homogeniczne)

Random Forest Regression

