



Politechnika Wrocławska

# Uczenie maszynowe jako narzędzie doskonalenia procesów wytwórczych

Mateusz Tabor

1

Wstep

## Przegląd literatury

# 3

## Metodyka

# Charakterystyka algorytmów uczenia maszynowego

## 4.1 Uczenie nadzorowane (Supervised Learning)

Uczenie nadzorowane to technika, w której algorytm otrzymuje gotowy zestaw danych z wcześniej określonymi etykietami i uczy się na ich podstawie przewidywać etykiety dla nowych nieznanymi danych. Dane uczące zawierają zmienne wejściowe oraz zmienną predykowaną (etykietę). Standardowy proces uczenia nadzorowanego obejmuje podział zbioru danych na zbiór treningowy, walidacyjny oraz testowy. Trening polega na dopasowaniu parametrów modelu do danych treningowych poprzez minimalizację funkcji straty, która mierzy różnicę między przewidywaniami modelu a rzeczywistymi etykietami. Po zakończeniu treningu danych model zostaje poddany ocenie na zbiorze walidacyjnym w celu doboru hiperparametrów oraz monitorowaniu uczenia maszynowego, aby zapobiec przeuczeniu modelu. Ostateczna ocena odbywa się na zbiorze testowym na danych nieznanymi dla modelu i na podstawie tego modelu określana wartość dokładności, precyzji, recall, F1 (dla klasyfikacji) lub MSE, MAE (dla regresji) (Bishop, 2006; Hastie et al., 2009).

### 4.1.1 Regresja liniowa (Linear Regression)

Regresja liniowa to podstawowa technika statystyczna i uczenia maszynowego stosowana do modelowania zależności między zmienną zależną (predykowaną) a jedną lub więcej zmiennymi niezależnymi (cechami). Celem regresji liniowej jest znalezienie liniowej funkcji, która najlepiej opisuje związek między zmiennymi, umożliwiając przewidywanie wartości zmiennej zależnej na podstawie wartości zmiennych niezależnych.

**Regresja liniowa prosta.** W przypadku pojedynczej zmiennej niezależnej model można opisać równaniem:

$$y = \alpha + \beta x + \varepsilon, \quad (4.1)$$

gdzie  $y$  to zmienna zależna,  $x$  to zmienna niezależna,  $\alpha$  to wyraz wolny,  $\beta$  to współczynnik nachylenia linii regresji, a  $\varepsilon$  to składnik losowy (błąd modelu).

**Regresja wieloraka.** Regresja wieloraka jest rozszerzeniem regresji prostej, poprzez używanie wielu zmiennych niezależnych. Zakłada się liniową zależność

między zmienną zależną a kombinacją liniową zmiennych niezależnych:

$$y = \alpha + \sum_{i=1}^p \beta_i x_i + \varepsilon, \quad (4.2)$$

gdzie  $y$  to zmienna zależna,  $\alpha$  to wyraz wolny,  $x_i$  to zmienne niezależne,  $\beta_i$  to współczynniki regresji określające wpływ danej zmiennej na zmienną zależną, a  $\varepsilon$  to składnik losowy (błąd modelu).

Celem algorytmu regresji (prostej i wielorakiej) jest znalezienie optymalnych wartości współczynników  $\beta_i$ , które minimalizują błąd predykcji. Współczynniki można dobrać za pomocą różnych metod, jednak najczęściej stosuje się metodę najmniejszych kwadratów (OLS — Ordinary Least Squares), która minimalizuje sumę kwadratów różnic między rzeczywistymi a przewidywanymi wartościami zmiennej zależnej.

**Estymator OLS (macierzowy zapis).**

$$\hat{\beta} = (X^\top X)^{-1} X^\top \mathbf{y} \quad (4.3)$$

Wzór (4.3) to macierzowy zapis estymatora OLS. Wyjaśnienie:

- $X$  — macierz projektująca (design matrix) o wymiarach  $n \times p$  (lub  $n \times (p+1)$ , jeśli dodano kolumnę jedynek dla wyrazu wolnego),
- $\mathbf{y}$  — wektor obserwacji o wymiarze  $n \times 1$ ,
- $\hat{\beta}$  — wektor estymowanych współczynników o wymiarze  $p \times 1$ .

Aby wyrażenie było poprawne, macierz  $X^\top X$  musi być odwracalna (brak doskonałej multikolinearności). Przy klasycznych założeniach (m.in.  $E[\varepsilon] = 0$ ,  $\text{Var}(\varepsilon) = \sigma^2 I$ ) estymator jest nieobciążony, a jego wariancja wynosi  $\text{Var}(\hat{\beta}) = \sigma^2 (X^\top X)^{-1}$ . Gdy  $X^\top X$  jest źle uwarunkowana lub nieodwracalna, stosuje się regularyzację (np. Ridge) lub metody numeryczne (gradient descent, SVD) (James et al., 2013).

**Inne metody estymacji współczynników regresji liniowej:**

- Metoda gradientu prostego (Gradient Descent) — iteracyjna metoda optymalizacji minimalizująca funkcję straty poprzez aktualizację współczynników w kierunku przeciwnym do gradientu (Bishop, 2006; Murphy, 2012).
- Metoda najmniejszych modułów (Least Absolute Deviations) — minimalizuje sumę bezwzględnych różnic między rzeczywistymi a przewidywanymi wartościami, co czyni ją bardziej odporną na wartości odstające (Birkes and Dodge, 1993).
- Metoda Ridge Regression — wprowadza regularyzację L2, karę za duże wartości współczynników, co pomaga w radzeniu sobie z problemem multikolinearności i przeuczenia modelu (Hoerl and Kennard, 1970; Hastie et al., 2009).

- Metoda Lasso Regression — regularyzacja L1, która może prowadzić do zerowania niektórych współczynników, skutkując modelem o mniejszej liczbie cech (automatyczny wybór cech) (Tibshirani, 1996).
- Metoda Elastic Net — łączy regularyzację L1 i L2, co pozwala na lepsze dostosowanie modelu do danych (Zou and Hastie, 2005).
- Metoda SVD (Singular Value Decomposition) — rozkłada macierz projektującą na składniki, umożliwiając stabilne obliczenie współczynników regresji nawet w przypadku kolinearności cech (Golub and Loan, 1996; Press et al., 2007).
- Metoda Bayesian Regression — wykorzystuje podejście bayesowskie do estymacji współczynników, uwzględniając niepewność i priorytety w modelu (Gelman and Hill, 2006; Bishop, 2006).
- QR Decomposition — rozkłada macierz projektującą na iloczyn macierzy ortogonalnej i górnortrójkątnej, co umożliwia efektywne rozwiązanie układu równań regresji (Golub and Loan, 1996; Press et al., 2007).

#### 4.1.2 Regresja logistyczna (Logistic Regression)

Regresja logistyczna to technika statystyczna i uczenia maszynowego stosowana do modelowania zależności między zmienną zależną a jedną lub więcej zmiennymi niezależnymi, gdy zmienna zależna przyjmuje wartości binarne (np. 0 lub 1, tak lub nie). Celem regresji logistycznej jest przewidywanie prawdopodobieństwa przynależności do jednej z dwóch klas na podstawie wartości zmiennych niezależnych.

**Model regresji logistycznej.** Model regresji logistycznej można zapisać jako:

$$P(Y = 1|X) = \sigma(\beta^\top X) = \frac{1}{1 + e^{-\beta^\top X}} \quad (4.4)$$

gdzie:

- $P(Y = 1|X)$  — prawdopodobieństwo, że zmienna zależna  $Y$  przyjmuje wartość 1, biorąc pod uwagę zmienne niezależne  $X$ ,
- $\sigma(z)$  — funkcja sigmoidalna, która przekształca dowolną wartość rzeczywistą  $z$  w przedział  $(0, 1)$ .

**Estymacja parametrów.** Parametry modelu  $\beta$  są estymowane za pomocą metody największej wiarygodności (Maximum Likelihood Estimation, MLE). Celem jest maksymalizacja funkcji wiarygodności:

$$L(\beta) = \prod_{i=1}^n P(Y_i|X_i; \beta) \quad (4.5)$$

co jest równoważne minimalizacji funkcji straty:

$$J(\beta) = - \sum_{i=1}^n [Y_i \log(P(Y_i|X_i; \beta)) + (1 - Y_i) \log(1 - P(Y_i|X_i; \beta))] \quad (4.6)$$

**Właściwości modelu.** Model regresji logistycznej ma kilka istotnych właściwości:

- Wynikiem modelu jest prawdopodobieństwo, co czyni go odpowiednim narzędziem do klasyfikacji binarnej.
- Model jest odporny na wartości odstające, ponieważ wykorzystuje funkcję sigmoidalną.
- Można go łatwo rozszerzyć na problemy wieloklasowe (np. regresja wielomianowa).

#### 4.1.3 $k$ -Najbliższych Sąsiadów (k-Nearest Neighbors)

Algorytm  $k$ -Najbliższych Sąsiadów umieszcza dane wejściowe w przestrzeni wielowymiarowej i klasyfikuje je na podstawie etykiet najbliższych sąsiadów w tej przestrzeni. Przestrzeń jest definiowana przez cechy danych, zbiór danych posiadający  $x$  cech jest reprezentowany w  $x$ -wymiarowej przestrzeni. Algorytm klasyfikując dany obiekt oblicza odległości między nim a wszystkimi innymi obiektami w przestrzeni, a następnie wybiera  $k$  najbliższych sąsiadów. Wartość  $k$  jest ustalana przed rozpoczęciem działania algorytmu. Niska wartość parametru  $k$  jest bardziej podatna na szumy w danych, podczas gdy wysoka wartość  $k$  może prowadzić do nadmiernego uogólnienia modelu (Cover and Hart, 1967).

Algorytm  $k$ -najbliższych sąsiadów może wykorzystywać różne metryki do obliczania odległości m.in:

**Metryka Euklidesowa:** Najpowszechniejsza metryka używana do obliczania odległości między dwoma punktami w przestrzeni wielowymiarowej. Definiowana jest jako:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (4.7)$$

gdzie  $p$  i  $q$  to dwa punkty w przestrzeni, a  $n$  to liczba wymiarów. Wyliczanie odległości metryką euklidesową polega na policzeniu różnicy odległości w każdym wymiarze dla dwóch punktów, zsumowaniu kwadratów tych różnic, a następnie wyciągnięciu pierwiastka kwadratowego z tej sumy.

**Metryka Manhattan:** Metryka Manhattan, nazywana również metryką taksówkową lub L1, mierzy odległość między dwoma punktami jako sumę wa-



tości bezwzględnych z różnic współrzędnych. Definiowana jest jako:

$$d(p, q) = \sum_{i=1}^n |p_i - q_i| \quad (4.8)$$

gdzie  $p$  i  $q$  to dwa punkty w przestrzeni, a  $n$  to liczba wymiarów. Obliczana jest różnica wartości  $p$  i  $q$  dla każdego wymiaru, a następnie sumowane są wartości bezwzględne tych różnic.

**Metryka Kosinusowa:**

5

# Zastosowania uczenia maszynowego w procesach wytwórczych

6

## Studium przypadku

7

## Wnioski końcowe

# Bibliografia

- David Birkes and Yadolah Dodge. *Alternative Methods of Regression*. Wiley, New York, 1993.
- Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 2006.
- Thomas M. Cover and Peter E. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967. doi: 10.1109/TIT.1967.1053964.
- Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge University Press, Cambridge, 2006.
- Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 3rd edition, 1996.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2nd edition, 2009.
- Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970. doi: 10.1080/00401706.1970.10488634.
- Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer, New York, 1st edition, 2013.
- Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA, 2012.
- William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 3rd edition, 2007.
- Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. doi: 10.1111/j.2517-6161.1996.tb02080.x.
- Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005. doi: 10.1111/j.1467-9868.2005.00503.x.