

WSI Zadanie 4 SVM

Piotr Lenczewski

13.12.2023

1 Opis badanego algorytmu

1.1 Cele algorytmu

Celem algorytmu SVM (Support Vector Machines) jest klasyfikacja. Szukamy funkcji decyzyjnej (klasyfikatora), który dla danego wektora x wskaże najbardziej prawdopodobną etykietę y (w naszym przypadku -1 lub 1).

1.2 Forma dualna

Algorytm SVM wykorzystany w tym sprawozdaniu posiada formę dualną, co umożliwia działanie algorytmu dla różnych jąder $K(u, v)$. W tym przypadku:

- wielomianowe $(1 + u \cdot v^T)^{\text{degree}}$
- rbf (Radial Basis Function) $\exp\left(-\frac{\|u-v\|^2}{2\sigma^2}\right)$

gdzie degree to stopień wielomianu, a σ kontroluje "szerokość" jądra.

Pozwala to na efektywną implementację w przypadku dużych zbiorów danych.

1.3 Funkcja decyzyjna

Dla formy dualnej funkcję decyzyjną formułuje się jako:

$$y = \text{sign} \left(\sum_{i, \alpha_i \geq 0} \alpha_i y_i K(x_i, z_i) + b \right)$$

Gdzie:

- α_i to współczynniki Lagrange'a związane z danym zbiorem treningowym,
- y_i to etykiety dla wektorów treningowych,
- x_i to zbiór wektorów treningowych,
- z_i to zbiór wektorów testowych,
- b to przesunięcie.

1.4 Optymalizacja funkcji decyzyjnej

Problem optymalizacji funkcji decyzyjnej można zapisać jako:

$$\max_{\alpha} \mathcal{L}_{\text{dual}}(\alpha) = \max_{\alpha} \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right\}$$

Do optymalizacji używamy metody gradientu. Dla tak zdefiniowanego problemu należy optymalizować α_i oraz b .

1.5 Optymalizacja α

Gradient dla α_i można policzyć ze wzoru:

$$\nabla a_k = 1 - y_k \sum_{j=1}^n \alpha_j y_j K(x_j, x_k)$$

Zatem α optymalizujemy:

$$a_k = a_k + \text{lr} \cdot \nabla a_k$$

Gdzie lr to learning rate, ustawiony na najlepszą znalezioną wartość dla danych hiperparametrów.

1.6 Optymalizacja b

Po zoptymalizowaniu α_i , b można policzyć ze wzoru:

$$b = \text{avg}_{i, C \leq \alpha_i \leq 0} \left\{ y_i - \sum_{j, \alpha_j \geq 0} \alpha_j y_j K(x_j, x_i) \right\}$$

1.7 Parametr C

Parametr C jest hiperparametrem, który reguluje kompromis między uzyskaniem jak najmniejszego błędu treningowego a minimalizacją nadmiernego dopasowania (overfitting). Ogranicza on współczynniki Lagrange'a: $0 \leq \alpha_i \leq C$

2 Planowane eksperymenty numeryczne

Mam zamiar zbadać wpływ rodzaju jądra i hiperparametrów na działanie algorytmu SVM. Będę operować na zbiorze danych (ograniczonym do 4000 elementów ze względów optymalizacyjnych) Wine Quality Data Set ze strony: <https://archive.ics.uci.edu/ml/datasets/wine+quality>

2.1 Założenia początkowe

- Wynik działania algorytmu SVM będę oceniać poprzez prawdopodobieństwo, że algorytm poprawnie przewidział etykietę dla testowych wektorów.
- Początkowo ustawiam hiperparametr C=1.
- Parametr learning rate ustawiam po kolei na wartości najkorzystniejsze dla konkretnych wartości hiperparametrów i sprawdzam te wartości dla wszystkich badanych hiperparametrów.

2.2 Jądro wielomianowe

Zamierzam zbadać wynik działania dla jądra wielomianowego dla kilku prób dla różnych wartości hiperparametru degree.

2.3 Jądro rbf

Zamierzam zbadać wynik działania dla jądra rbf dla kilku prób dla różnych wartości hiperparametru σ .

2.4 Hiperparametr C

Zamierzam zbadać wpływ hiperparametru C na wyniki działania dla pozostałych hiperparametrów ustawionych na najbardziej optymalne z dotychczas uzyskanych wartości.

3 Uzyskane wyniki

3.1 Jądro wielomianowe

Wynik działania algorytmu w dużej mierze zależy od parametru learning rate. Dla każdej wartości degree przy optymalnej dla niej wartości learning rate wynik działania algorytmu jest podobny.

learning rate	degree = 1	degree = 2	degree = 3
1e-4	0.709	0.479	0.636
1e-6	0.7	0.721	0.609
1e-8	0.664	0.611	0.7

Table 1: Wyniki działania SVM dla jądra wielomianowego dla różnych degree

3.2 Jądro rbf

Zależności pomiędzy wynikami działania algorytmu dla różnych wartości σ są podobne dla różnych wartości learning rate.

learning rate	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 10$
1e-2	0.58	0.78	0.728
1e-3	0.59	0.791	0.718
1e-4	0.708	0.746	0.715

Table 2: Wyniki działania SVM dla jądra rbf dla różnych σ

3.3 Hiperparametr C

Przy badaniu wpływu hiperparametru C na działanie algorytmu będę korzystał z jądra rbf i hiperparametru $\sigma = 1$

Próba	C = 0.1	C = 1	C = 10	C = 100
1.	0.755	0.789	0.803	0.794
2.	0.755	0.793	0.793	0.788
3.	0.753	0.789	0.793	0.793

Table 3: Wyniki działania SVM dla różnych wartości C

4 Wnioski

- Parametr learning rate przyjmuje różne wartości optymalne dla różnych wartości hiperparametrów.
- Jądro rbf przy odpowiednich wartościach parametrów jest bardziej optymalne od jądra wielomianowego.
- Dla jądra wielomianowego osiągnane wyniki dla różnych wartości degree były podobne
- Dla jądra rbf widać, że optymalna wartość parametru σ jest gdzieś w pobliżu 1
- Optymalna wartość hiperparametru C jest gdzieś w pobliżu wartości 10