

Optymalizacja hiperparametrów w głębokich sieciach neuronowych

Jacek Tyszkiewicz

Opiekun: dr hab. inż. Wojciech Chmiel

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

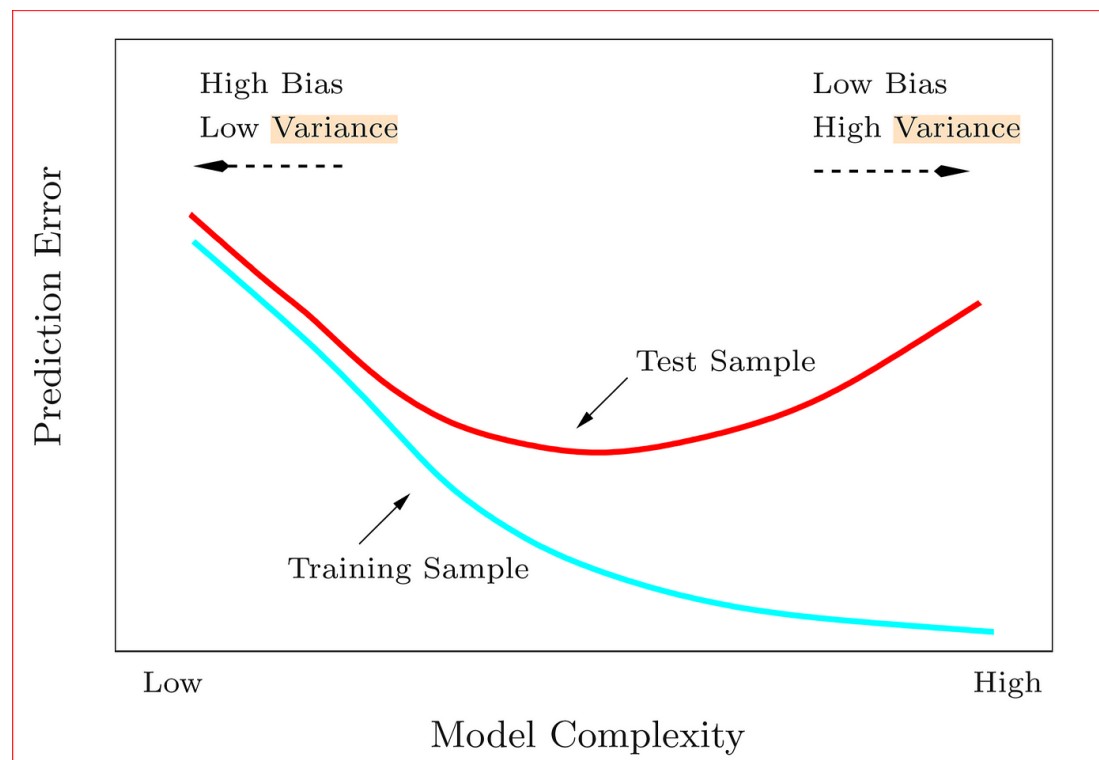
AGH University of Krakow



Plan prezentacji

1. Wprowadzenie do problematyki optymalizacji hiperparametrycznej
2. Omówienie funkcji kosztu optymalizacji hiperparametrycznej
3. Przestrzeń poszukiwań
4. Zapoznanie z powszechnie uznanymi algorytmami optymalizacji hiperparametrycznej
5. Omówienie autorskiego algorytmu
6. Prezentacja wyników
7. Podsumowanie

Zrozumienie problematyki tworzenia architektury sieci neuronowej



Funkcja celu optymalizacji hiperparametrycznej

$$\lambda^* \approx \arg \min_{\lambda \in \Lambda} \text{mean}_{x \in X^{(\text{valid})}} \mathcal{L}(x; \mathcal{A}_\lambda(X^{(\text{train})})) \quad (1)$$

$$= \arg \min_{\lambda \in \Lambda} \Psi(\lambda) \quad (2)$$

$$\approx \arg \min_{\lambda \in \{\lambda^{(1)}, \dots, \lambda^{(S)}\}} \Psi(\lambda) \equiv \bar{\lambda} \quad (3)$$

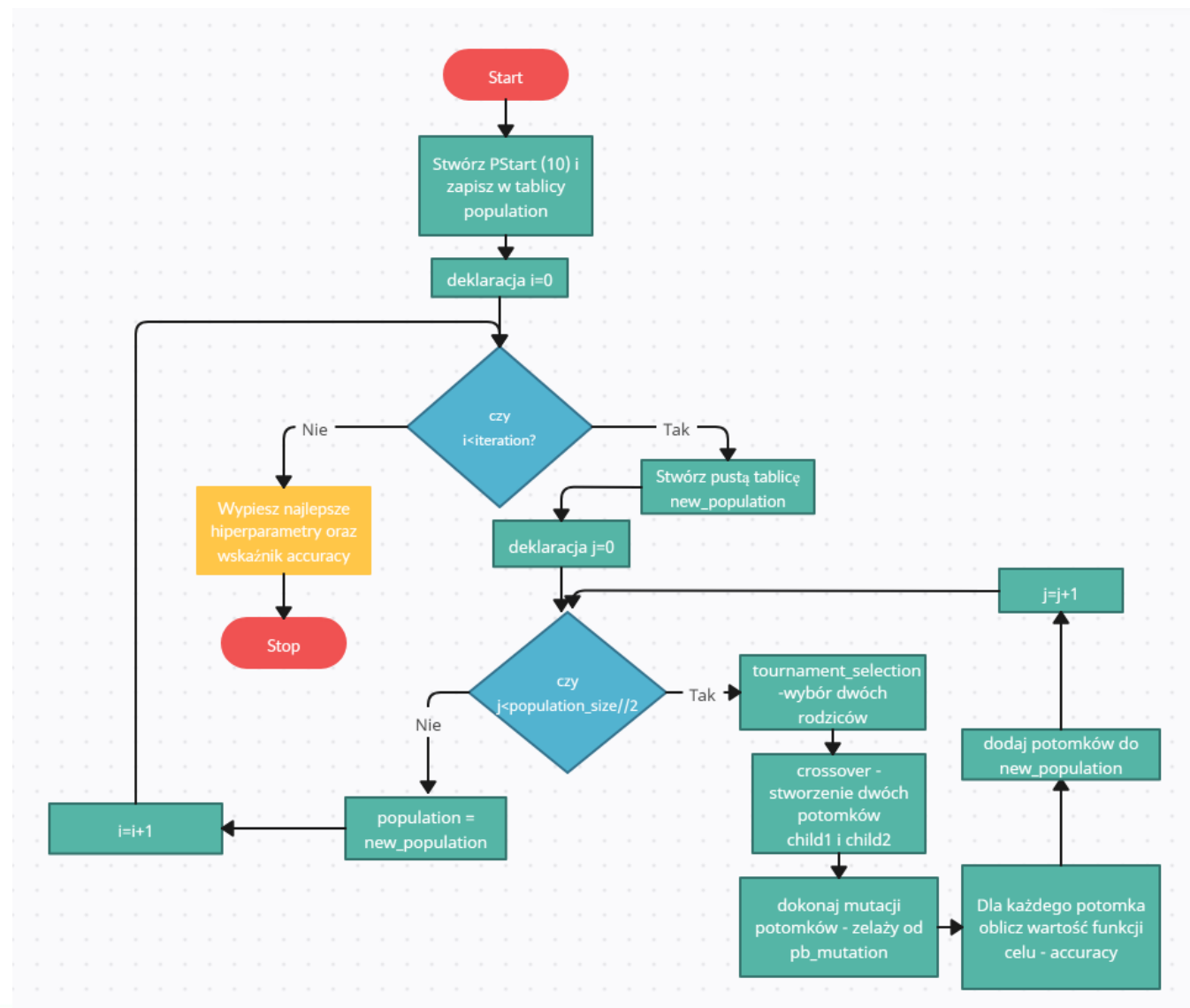
Przestrzeń poszukiwań

- Współczynnik uczenia
- Liczba warstw gęstych
- Liczba neuronów w warstwie gęstej
- Liczba warstw konwolucyjnych
- Rozmiar kernelu
- Liczba filtrów
- Funkcja aktywacji

Powszechnie uznane algorytmy optymalizacji hiperparametrycznej

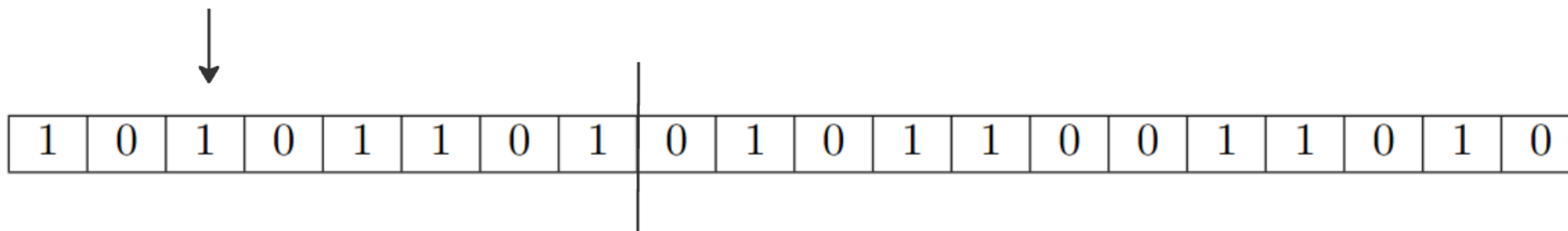
- Algorytmy Bayesowskie: BOGP, SMAC
- Algorytmy z grupy wyczerpujących poszukiwań: wyszukiwanie ręczne, wyszukiwanie siatką, wyszukiwanie losowe
- Algorytmy heurystyczne: Algorytm genetyczny, symulowanego wyżarzania

EA w wersji podstawowej, autorski algorytm optymalizacji hiperparametrycznej

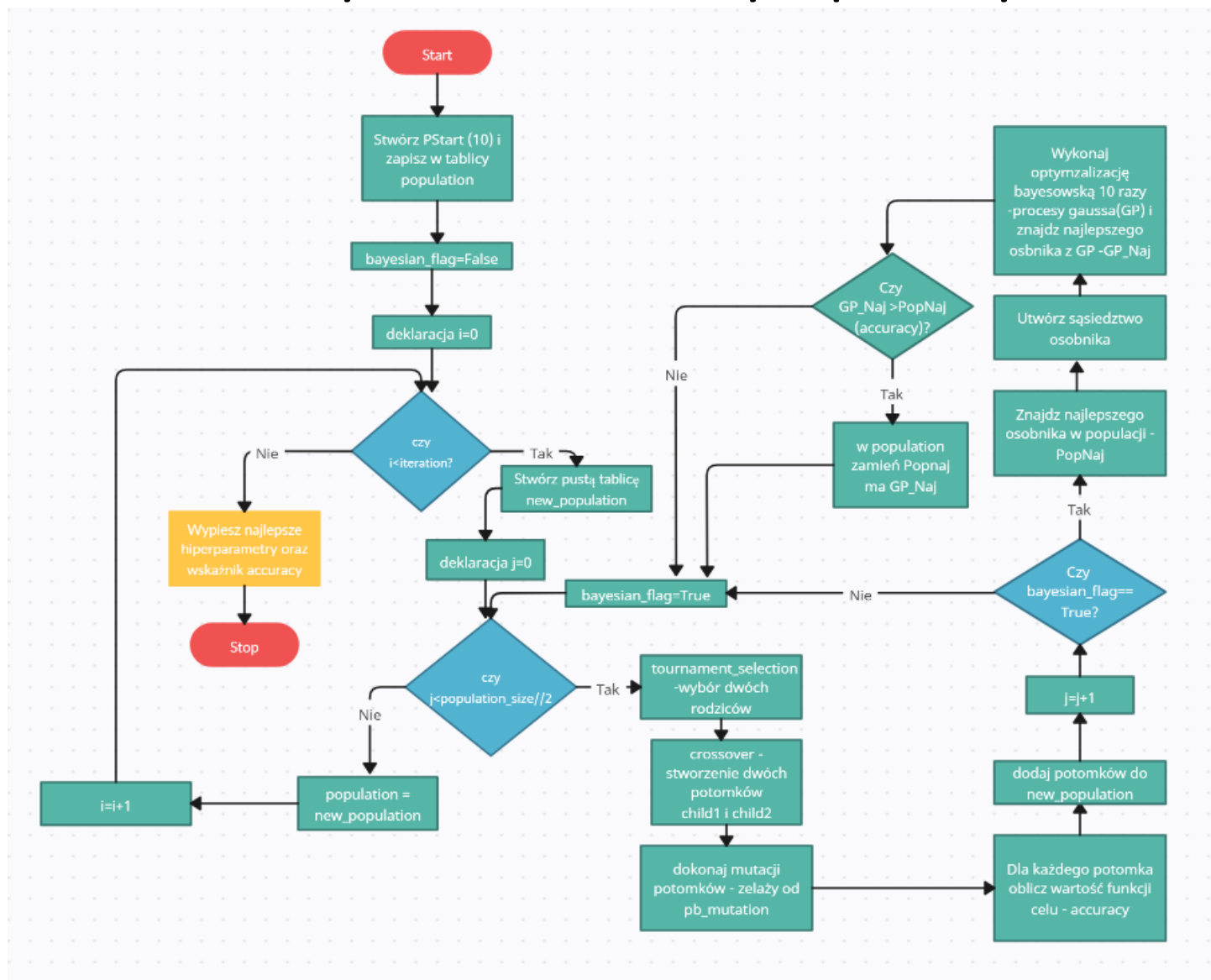


Najważniejsze komponenty EA

- Selekcja turniejowa
- krzyżowanie
- mutacja



BOinEA, czyli EA rozszerzony o procesy Gaussa



Wyniki na zbiorze Fashion-MNIST

Algorytm	Współczynnik uczenia	Liczba warstw gestych	Liczba neuronów w warstwie gestej	Funkcja aktywacyjna	Accuracy zbior treningowy [%]	Accuracy zbior testowy [%]	Czas obliczeń (sekundy)
BOGP	0.000995	1	297	ReLU	92.37	91.83	7304
SMAC	0.000395	2	418	ReLU	92.17	91.20	6512
EA	0.001852	2	274	ReLU	92.21	91.72	6424
BOinEA	0.002700	1	330	ReLU	92.16	91.57	6099

Table 3: Wyniki optymalizacji hiperparametrycznej

Wyniki na zbiorze NSL-KDD

Algorytm	Współczynnik uczenia	Liczba warstw gestych	Liczba neuronów w warstwie gestej	Funkcja aktywacyjna
BOGP	0.001521	2	80	ReLU
SMAC	0.003499	2	74	tanh
EA	0.008572	1	491	sigmoid
BOinEA	0.007011	1	361	ReLU

Table 4: Wyniki optymalizacji hiperparametrycznej (cz.1).

Algorytm	Dropout rate	Accuracy zbiorów treningowy [%]	Accuracy zbiorów testowy [%]	Czas obliczeń (sekundy)
BOGP	0	93.00	69.92	8595
SMAC	0	90.05	66.15	8363
EA	0.0584	94.22	70.17	7168.78
BOinEA	0.0557	92.83	69.56	6546

Table 5: Wyniki optymalizacji hiperparametrycznej (cz.2).

Wyniki na zbiorze CIFAR-10

Algorytm	Współczynnik uczenia	Liczba warstw gestych	Liczba neuronów w warstwie gestej	Liczba warstw konwolucyjnych	Rozmiar kernela
BOGP	0.000090	1	500	2	3x3
SMAC	0.000652	5	494	2	3x3
EA	0.001380	1	343	4	3x3
BOinEA	0.001834	3	446	3	3x3

Table 1: Wyniki optymalizacji hiperparametrycznej (cz.1).

Algorytm	Liczba filtrów	Funkcja aktywacyjna	Accuracy zbiorów treningowy [%]	Accuracy zbiorów testowy [%]	Czas obliczeń (sekundy)
BOGP	128	ReLU	74.36	73.68	8674
SMAC	128	ReLU	76.72	74.97	31208
EA	128	ReLU	78.62	78.22	33041
BOinEA	32	ReLU	77.28	75.00	26262

Table 2: Wyniki optymalizacji hiperparametrycznej (cz.2).

Podsumowanie

Zrealizowane cele:

- zapoznanie z podstawowymi algorytmami optymalizacji hiperparametrycznej,
- nauka tworzenia i doskonalenia architektury sieci neuronowej przy użyciu bibliotek hyperopt, scikit-optimize
- opracowanie autorskiego algorytmu optymalizacji hiperparparametrycznej

Kierunki rozwoju:

- rozwój w kierunku automatycznych systemów uczenia maszynowego