**Polecenie:**

Zapoznaj się z podręcznikiem - The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition - dostępnym tutaj: <https://link.springer.com/book/10.1007/978-0-387-84858-7>. Wyszukaj w nim informacji na temat bias - variance trade off.

Zapoznaj się z interaktywnym narzędziem to wytłumaczenia tej zależności: <https://mlu-explain.github.io/bias-variance/>

Na podstawie powyższych materiałów postaraj się krótko odpowiedzieć na następujące pytania:

1. Czym jest bias w uczeniu maszynowym?
2. Jaka jest zasada bias-variance tradeoff?
3. Jakie są przykłady modeli o dużym biasie?
4. Jakie są przykłady modeli o dużej wariancji?
5. W jaki sposób można zminimalizować bias w modelach?
6. Jakie techniki pomagają zmniejszyć wariancję

Materiały dodatkowe:  
<https://www.sciencedirect.com/science/article/pii/S2352012422009018>

1. **Czym jest bias w uczeniu maszynowym?**

Bias (błąd systematyczny) to różnica między oczekiwanymi (przeciętnymi) przewidywaniami modelu a rzeczywistymi wartościami, które próbujemy oszacować. Wysoki bias oznacza, że model jest zbyt uproszczony i nie jest w stanie uchwycić złożoności danych — prowadzi to do niedouczenia (underfitting).

1. **Jaka jest zasada bias-variance tradeoff?**

Bias-variance tradeoff polega na znalezieniu balansu pomiędzy zbyt prostym modelem (wysoki bias, niski variance), a zbyt złożonym modelem (niski bias, wysoki variance). Zbyt prosty model nie oddaje struktury danych (underfitting), a zbyt złożony dopasowuje się do szumów (overfitting). Celem jest minimalizacja całkowitego błędu poprzez kompromis między biasem a wariancją. Zazwyczaj zwiększając złożoność modelu, zmniejszamy bias, ale zwiększamy wariancję, i odwrotnie.

1. **Jakie są przykłady modeli o dużym biasie?**

Przykłady modeli o dużym biasie to:

* Regresja liniowa dla danych nieliniowych (model zbyt prosty względem danych)
* Drzewa decyzyjne o małej głębokości (np. stump — jedno rozgałęzienie)
* Naive Bayes (założenia niezależności często nie są spełnione, więc bias jest wysoki)

Te modele są bardzo uproszczone i nie są w stanie dobrze odwzorować skomplikowanych zależności w danych.

1. **Jakie są przykłady modeli o dużej wariancji?**

Przykłady modeli o dużej wariancji to:

* Drzewa decyzyjne o dużej głębokości (głęboko rozbudowane drzewa łatwo dopasowują się do szumu)
* K-nearest neighbors z małym k (np. k=1 — model "uczy się na pamięć" dane treningowe)
* Złożone modele nieliniowe (np. sieci neuronowe bez regularizacji)

Takie modele są bardzo elastyczne i wrażliwe na małe zmiany w danych treningowych (przeuczenie/overfitting).

1. **W jaki sposób można zminimalizować bias w modelach?**

Aby zmniejszyć bias, można:

* Używać bardziej złożonych modeli (np. przejście z regresji liniowej do wielomianowej)
* Dodawać nowe cechy (feature engineering)
* Stosować mniej restrykcyjne założenia względem danych
* Dobierać parametry tak, by model lepiej uczył się z danych (np. zwiększyć stopień wielomianu, głębokość drzewa)

Zwiększenie złożoności modelu zmniejsza bias, ale może podnieść wariancję.

1. **Jakie techniki pomagają zmniejszyć wariancję?**

Aby zmniejszyć wariancję modelu, stosuje się:

* Regularizację (np. Ridge, Lasso — ograniczanie złożoności modelu)
* Bagging (bootstrap aggregating — np. Random Forest)
* Zmniejszenie złożoności modelu (np. mniejsza głębokość drzewa, większe k w KNN)
* Cross-validation (do wyboru optymalnych parametrów)
* Early stopping (w sieciach neuronowych)

Te techniki pomagają ograniczyć "nadmierne dopasowanie" do danych treningowych i poprawić generalizację.