



Bias - variance trade off


1. Czym jest bias w uczeniu maszynowym?

-  **Bias (błąd systematyczny)** odnosi się do uproszczenia modelu, które prowadzi do niedouczenia (underfitting). To różnica między przewidywaniami modelu a rzeczywistymi wartościami, wynikająca z założeń przyjętych przez model. Model o wysokim bias nie jest w stanie dobrze uchwycić zależności w danych.


2. Jaka jest zasada bias-variance trade-off?

-  **Bias-variance trade-off** opisuje kompromis między błędem bias i błędem variance:
 - **wysoki bias, niskie variance** - model jest zbyt prosty i nie uczy się wystarczająco z danych
 - **niski bias, wysoki variance** - model jest zbyt skomplikowany i zbyt dokładnie dopasowuje się do danych uczących (overfitting)
- Celem jest znalezienie równowagi, w której całkowity błąd generalizacji (na nowych danych) jest jak najmniejszy.

3. Jakie są przykłady modeli o dużym biasie?

-  To takie modele, które mają ograniczoną złożoność i przez to upraszczają rzeczywistość:
 - regresja liniowa dla danych nieliniowych
 - płytkie drzewa decyzyjne
 - KNN z bardzo małą liczbą sąsiadów (np. $k = 1$) przy dużym szumie

4. Jakie są przykłady modeli o dużej wariancji?

-  Modele z dużą wariancją to te, które mają dużą elastyczność i mogą nadmiernie dopasować się do danych:
 - głębokie drzewa decyzyjne (bez przycinania)
 - KNN z bardzo małym k przy małej liczbie danych
 - sieci neuronowe bez rekuraryzacji i z małą ilością danych

5. W jaki sposób można zminimalizować bias w modelach?



Aby **zmniejszyć bias** można:

- używać bardziej złożone modele (np. zamiast regresji liniowej - sieci neuronowe lub modele nieliniowe)
- dodać nowe cechy
- zastosować transformację danych
- zapewnić odpowiednią ilość danych treningowych

6. Jakie techniki pomagają zmniejszyć wariancję?



Aby **zmniejszyć wariancję** można:

- zastosować rekularyzację (np. L1 lub L2)
- zwiększyć zbiór treningowy
- redukować liczbę cech
- użyć cross-validation do stabilnej oceny modelu
- użyć metody zespołowe np. bagging (np. Random Forests) lub boosting (np. Gradient Boosting Machines)