Data mining - zadanie 2

Gabriel Budziński

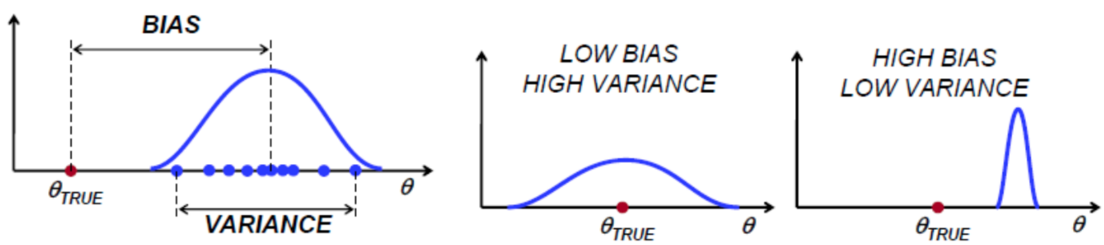
October 16, 2023

1 Treść

Recall what is the bias and variance of an estimator. Give examples of biased and unbiased estimators.

2 Rozwiązanie

- Bias - odległość wartości oczekiwanej estymatora od wartości parametru
- Variance - Odchylenia wartości estymatora od wartości oczekiwanej (po prostu jego wariancja)



- Bias estymatora $\bar{\theta}$ parametru θ zdefiniowany jest jako $bias(\bar{\theta}) = E[\bar{\theta}] - \theta$
- Estymator jest nieobciążony kiedy $bias(\bar{\theta}) = 0$, czyli $E[\bar{\theta}] = \theta$

Przykład estymatora nieobciążonego

Weźmy zmienne losowe $X_1, X_2, \ldots, X_n$ i zaproponujmy estymator wartości oczekiwanej μ:

$$\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Znajdźmy $E[\bar{\mu}]$:

$$E[\bar{\mu}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu$$

Skoro $E[\bar{\mu}] = \mu$, to $\bar{\mu}$ jest estymatorem nieobciążonym.

Przykład estymatora obciążonego

Zaproponujmy estymator wiaruancji zmiennych losowych, gdzie $\bar{\mu}$ jest nieobciążonym estymatorem wartości oczekiwanej.

$$\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\bar{\mu} + \bar{\mu}^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\bar{\mu}\overbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^2}^{\bar{\mu}} + \frac{1}{n}n\bar{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{\mu}^2$$

Z tego dalej, biorąc pod uwagę $\sigma^2 = E[X^2] - \mu^2$ oraz $Var[\bar{\mu}] = \frac{\sigma^2}{n}$

$$E[\bar{\sigma}^2] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{\mu}^2\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i^2] - E[\bar{\mu}^2] = \frac{1}{n}\sum_{i=1}^{n}(\sigma^2+\mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{(n-1)\sigma^2}{n}$$

Jak łatwo zauważyć, ten estymator ma bias $\frac{n-1}{n}$. Aby się go pozbyć, wystarczy przemnożyć go przez $\frac{n}{n-1}$