

Błąd selekcji wysokowymiarowego modelu liniowego

Jan Mielniczuk (IPI PAN, MiNI PW, miel@ipipan.waw.pl)

16 listopada 2012

Referat będzie kontynuacją referatu Piotra Pokarowskiego z 11.X.2012. Przedstawię w nim oszacowanie prawdopodobieństwa wyboru nadmodelu w trójstopniowej procedurze Screening-Ordering-Selection) \hat{M} selekcji modelu liniowego. Oszacowanie to, wraz z wynikami przedstawionymi poprzednio prowadzi do nierówności

$$P(\hat{M} \neq t) \leq e^{-\frac{r_n}{2}(1+o(1))},$$

gdzie t jest modelem prawdziwym o minimalnej liczbie atrybutów, a r_n karą związaną z dołączeniem jednego atrybutu w kryterium $GIC_j = n \log RSS_j + |j|r_n$. Omówię pewne zastosowania powyższej nierówności oraz porównam przyjęte założenia z założeniami potrzebnymi dla zgodności innych metod selekcji modelu liniowego rozpatrywanych w literaturze. Dla uzyskania wyniku wykorzystuje się nierówności dla prawego ogona rozkładu beta, które przedstawię.