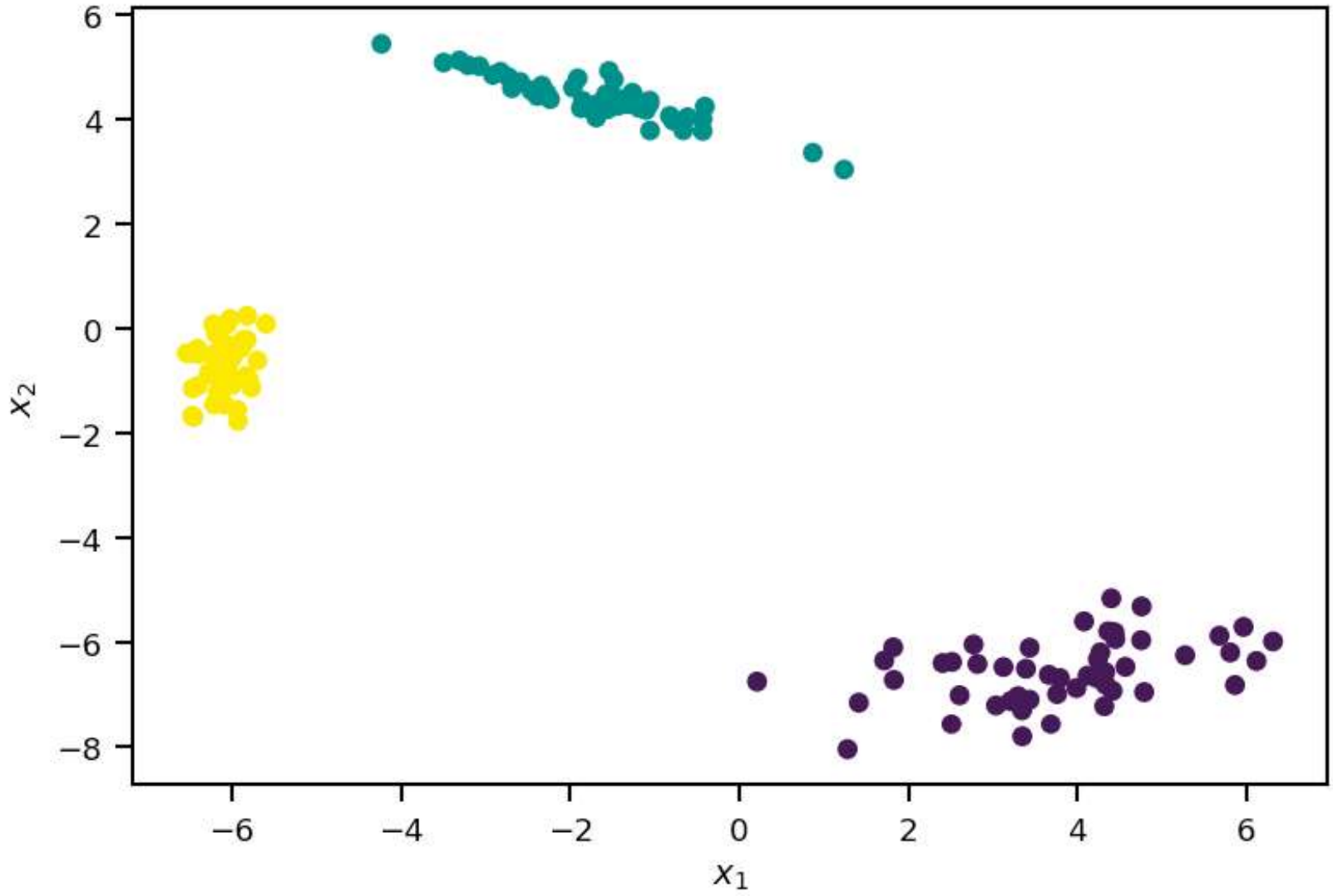
Lecture 17: Clustering and density estimation

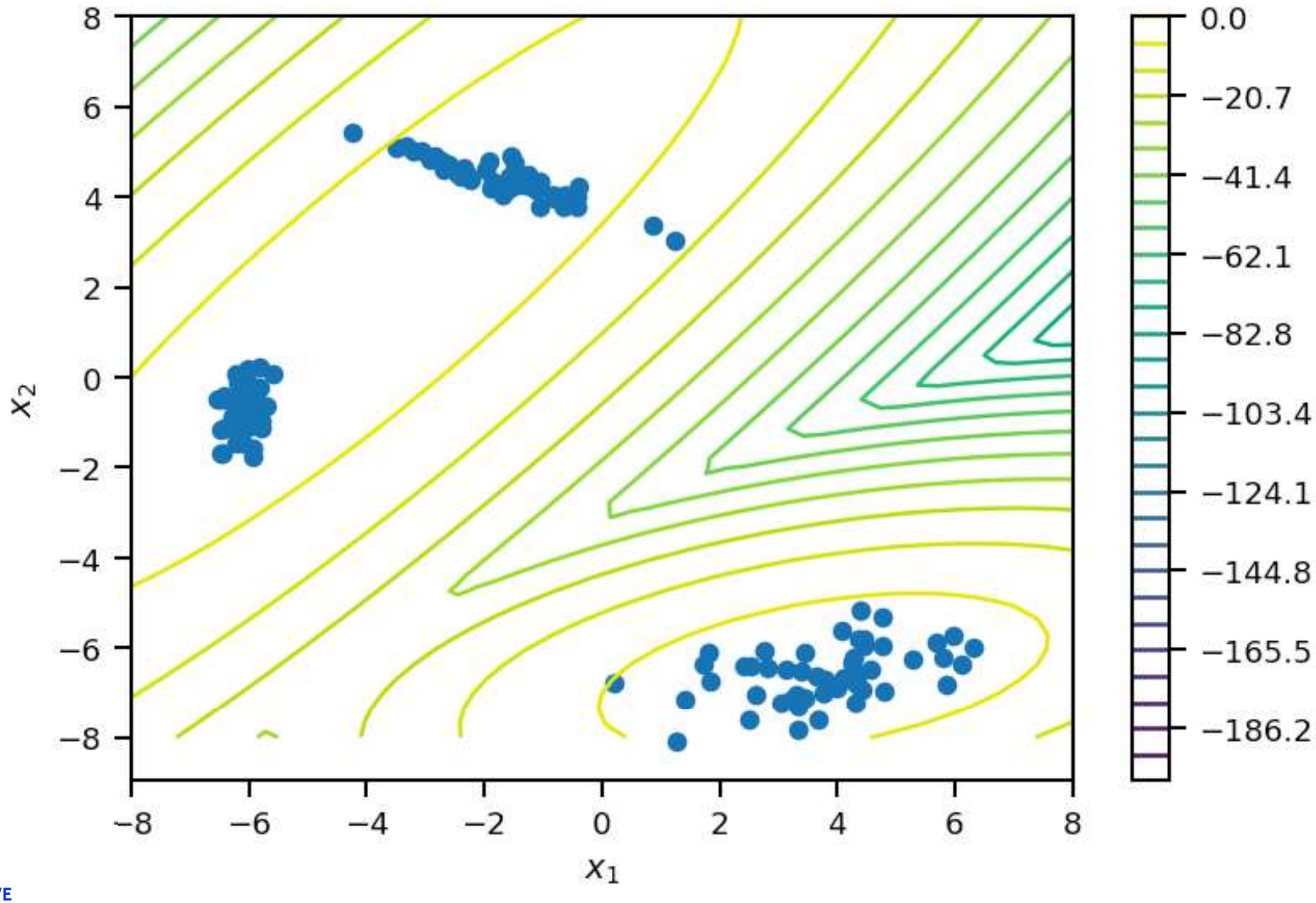
Professor Ilias Bilionis

Selecting the number of components using the Bayesian information criterion (BIC)

Example: Clustering with mixtures

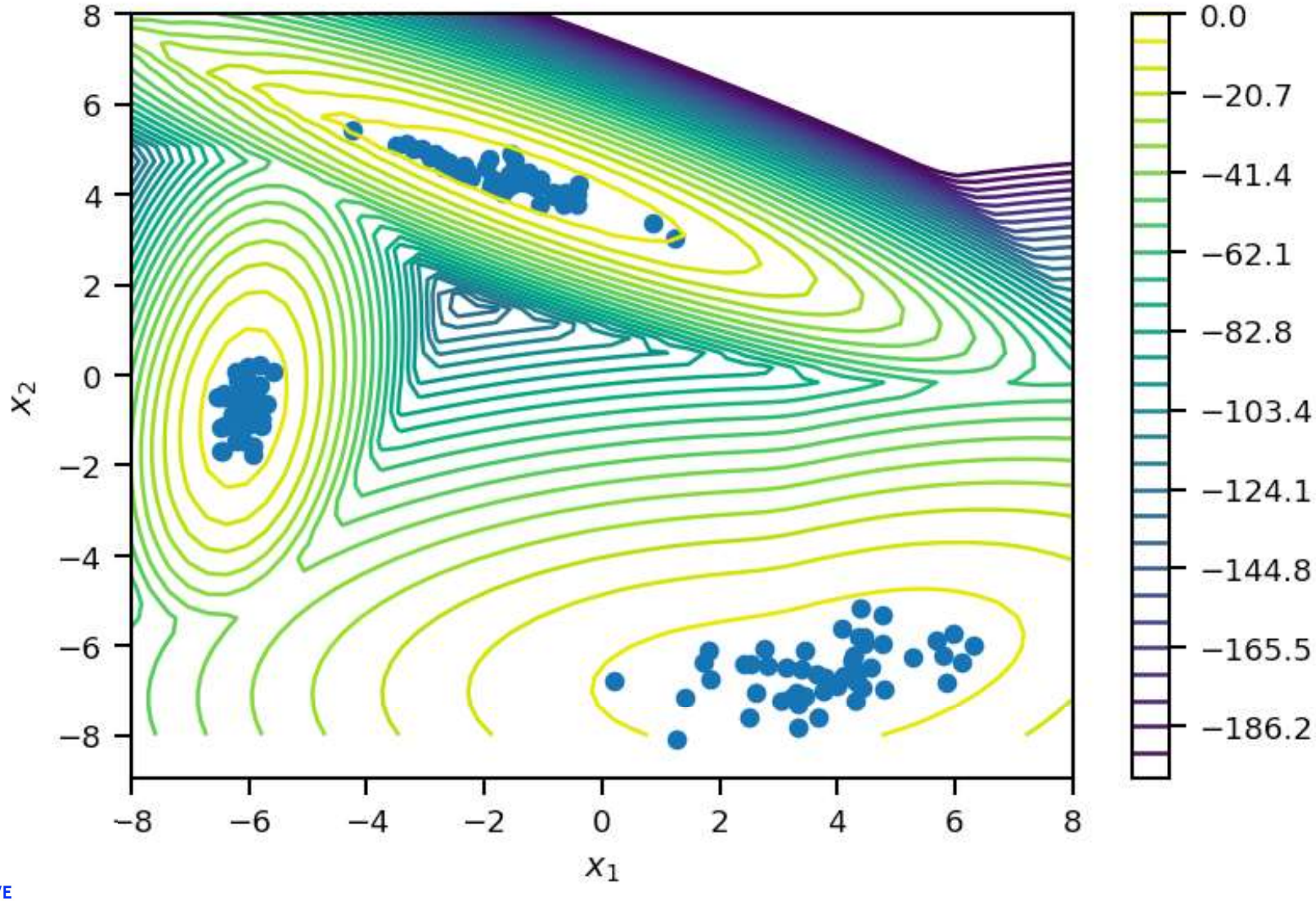


Trying fewer components doesn't work well



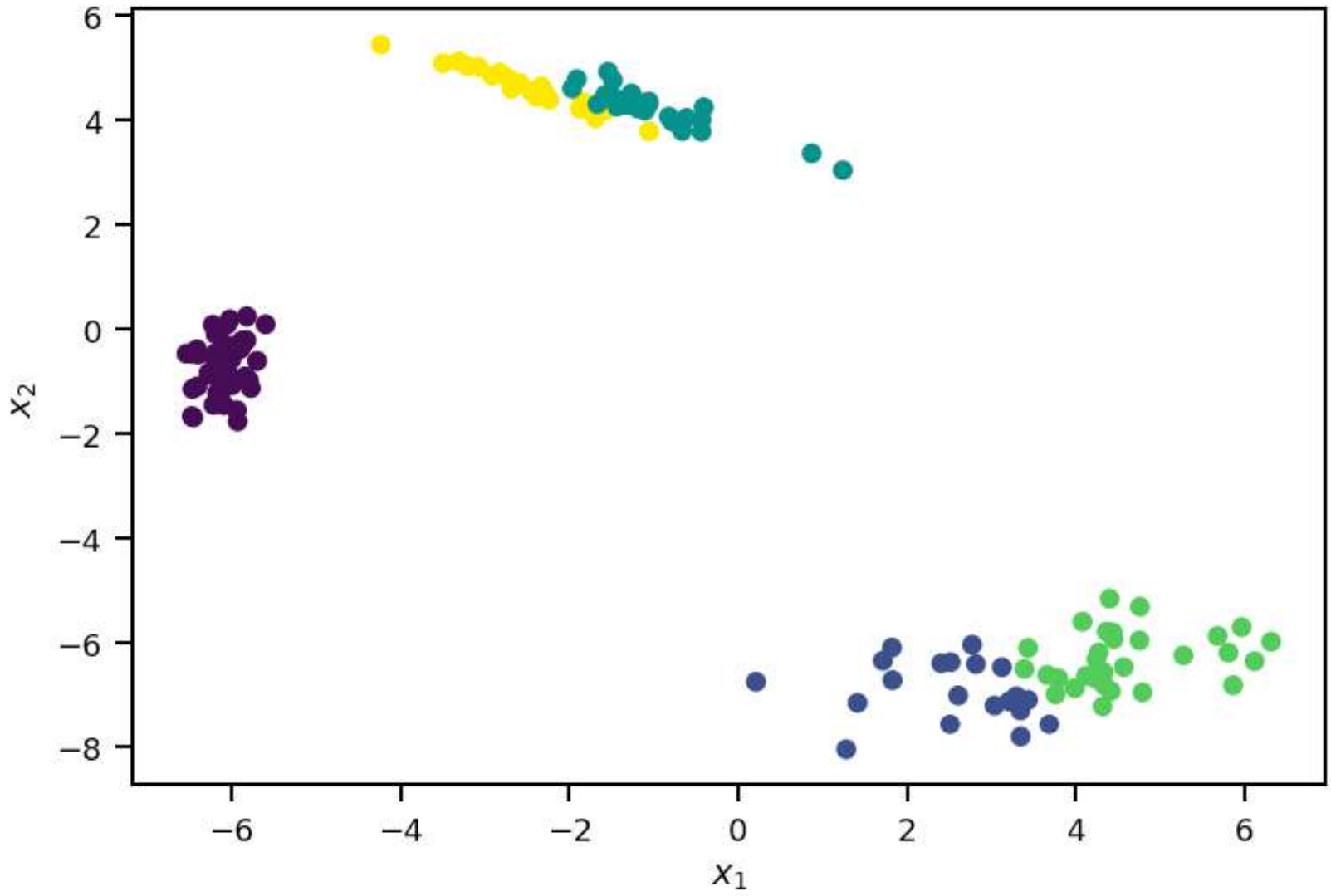
PREDICTIVE
SCIENCE LABORATORY

Trying more components works for density



PREDICTIVE
SCIENCE LABORATORY

But not for clustering



The Bayesian information criterion

$$BIC = \underbrace{\#\text{ of model parameters} \cdot \ln n}_{\text{Complexity}} \underbrace{- 2 \log\text{-likelihood}}_{\text{goodness-of-fit}}$$

best model: minimum BIC

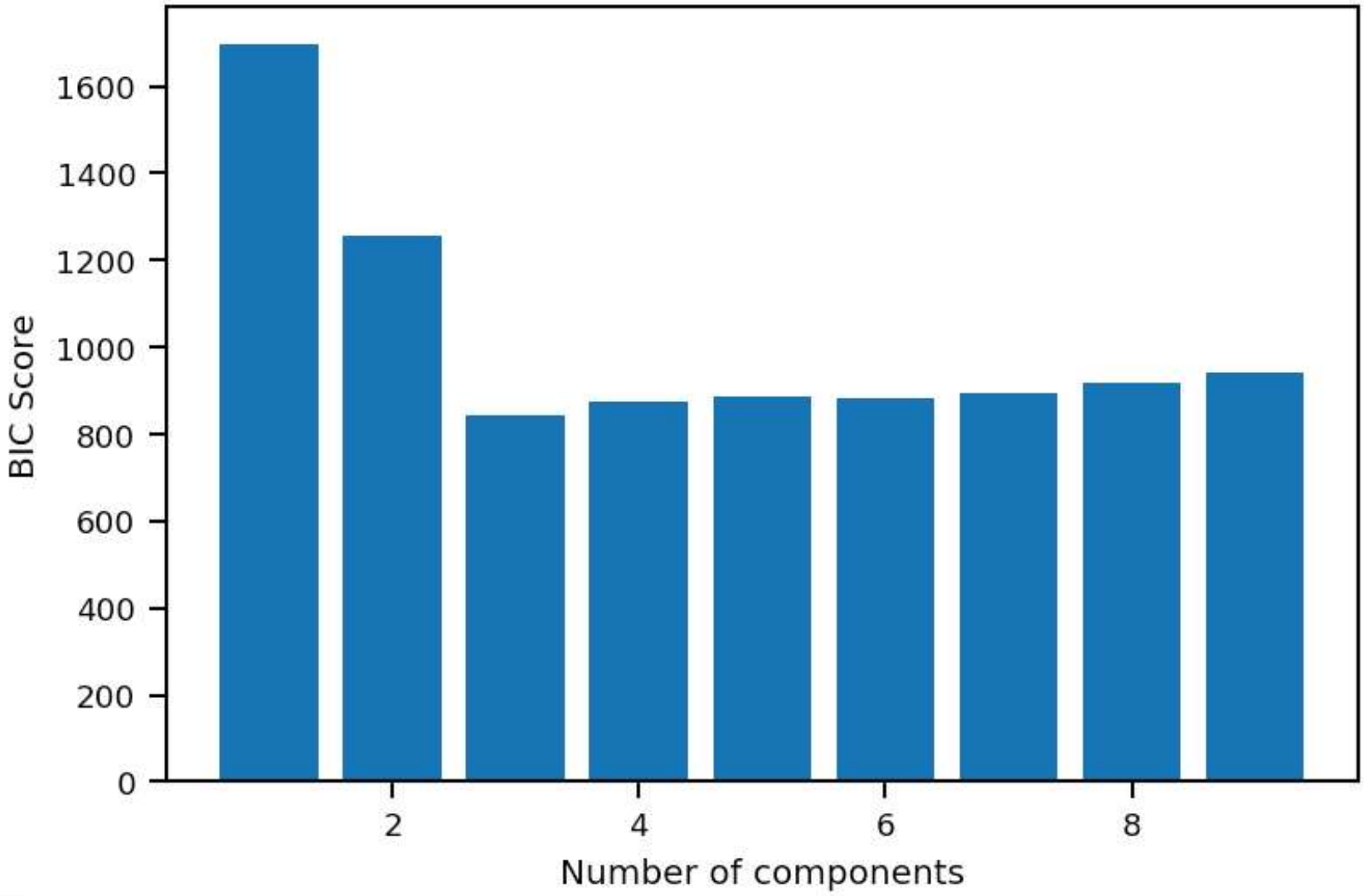
(Bayesian Model Section Lecture 28)

K clusters: # of model parameters $= K - 1 + K \cdot D + K \cdot D^2$ (I) (II) (III)

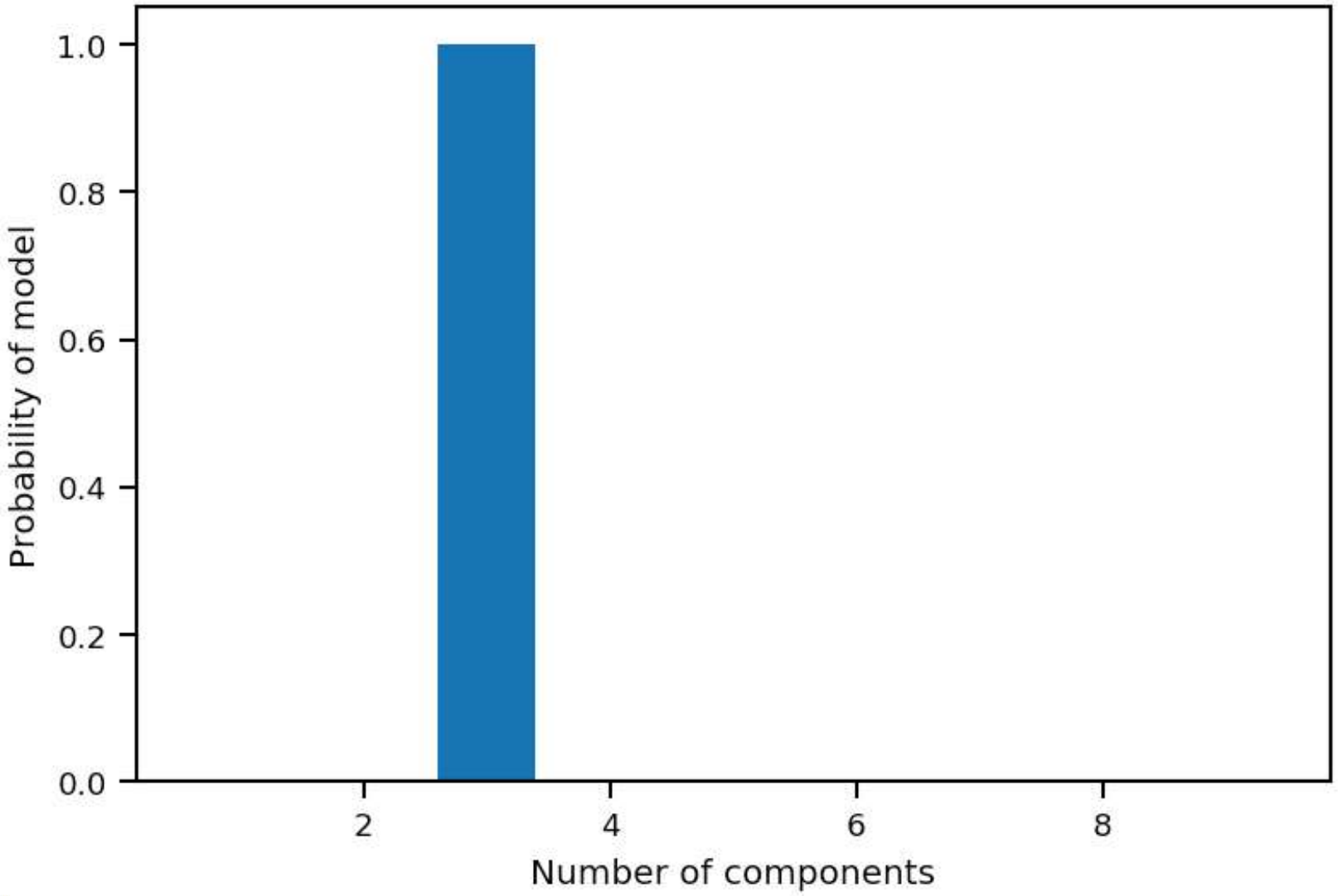
dimension of

$$BIC = \left[K - 1 + KD + KD^2\right] \ln n - 2 \log \text{likelihood}.$$

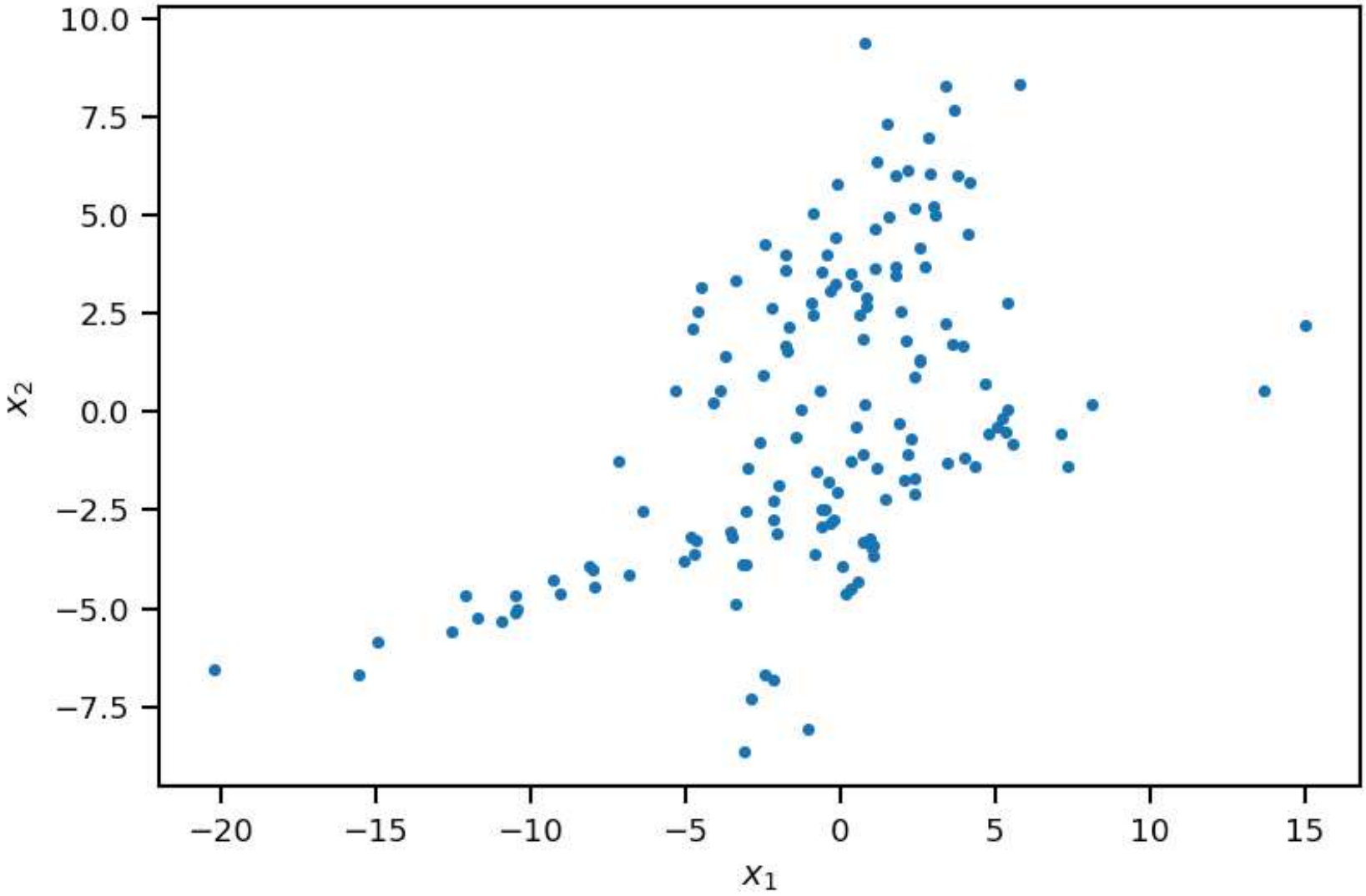
BIC scores on our original example



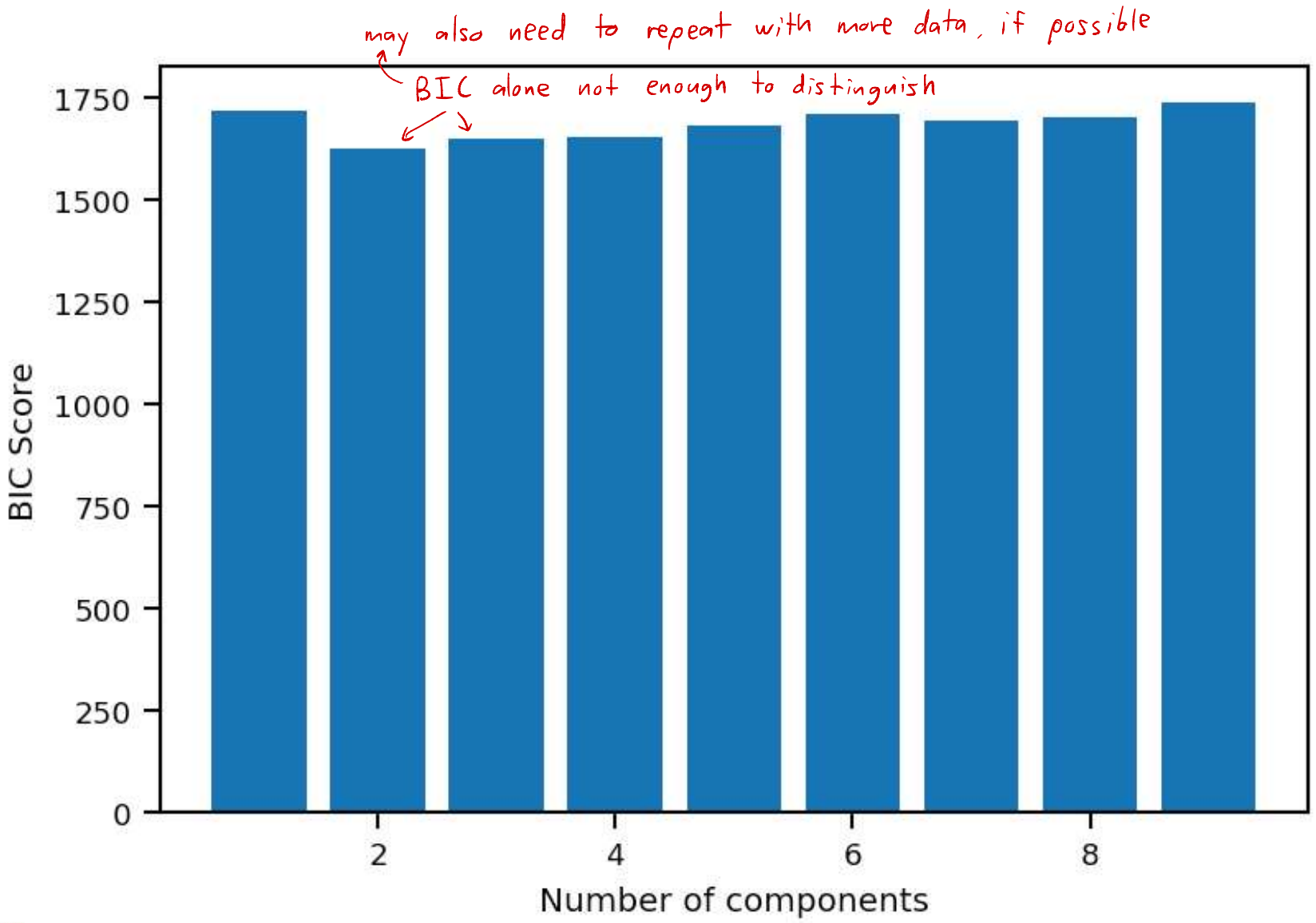
Probability over number of components



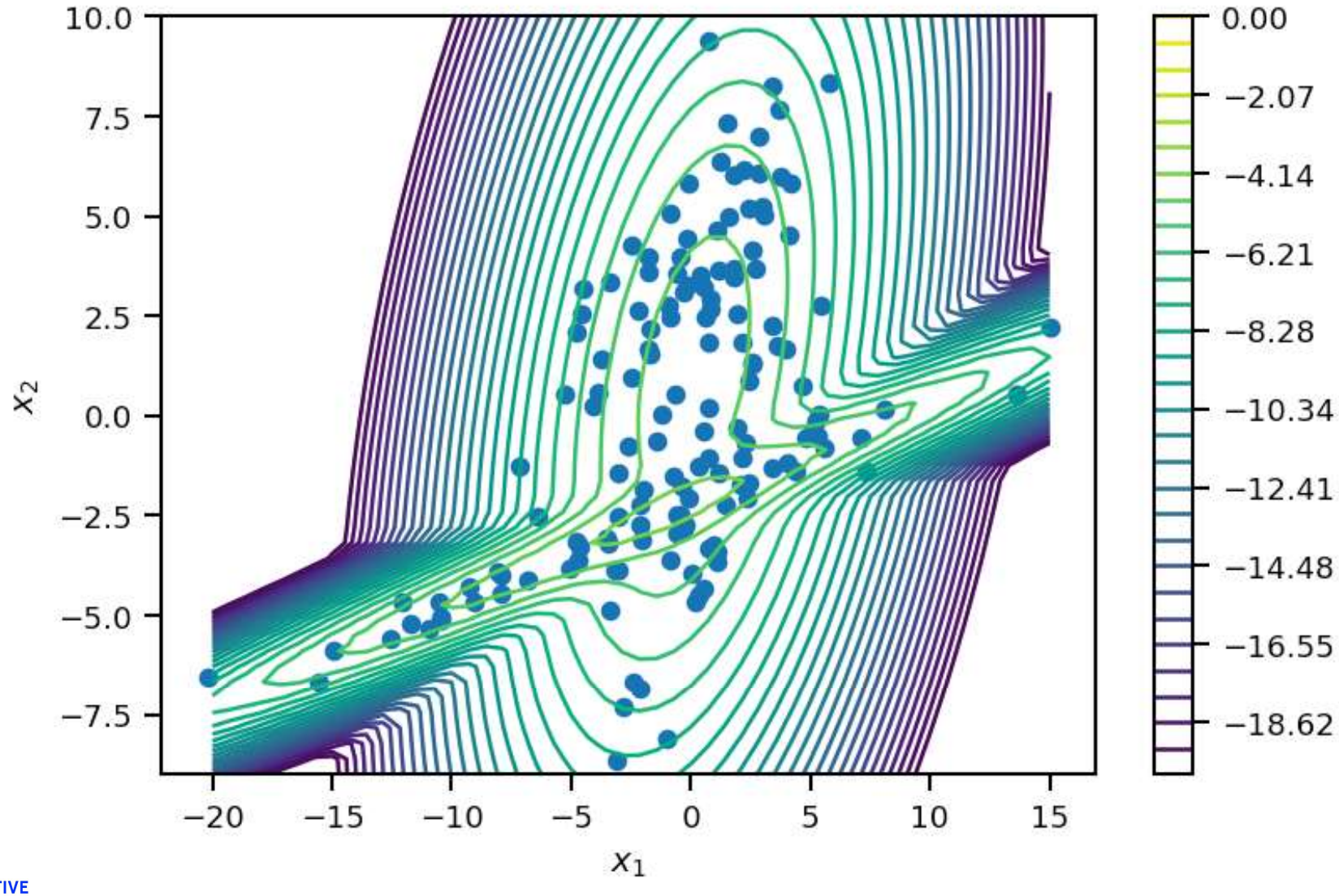
A less obvious example



A less obvious example



Most probable model



PREDICTIVE
SCIENCE LABORATORY