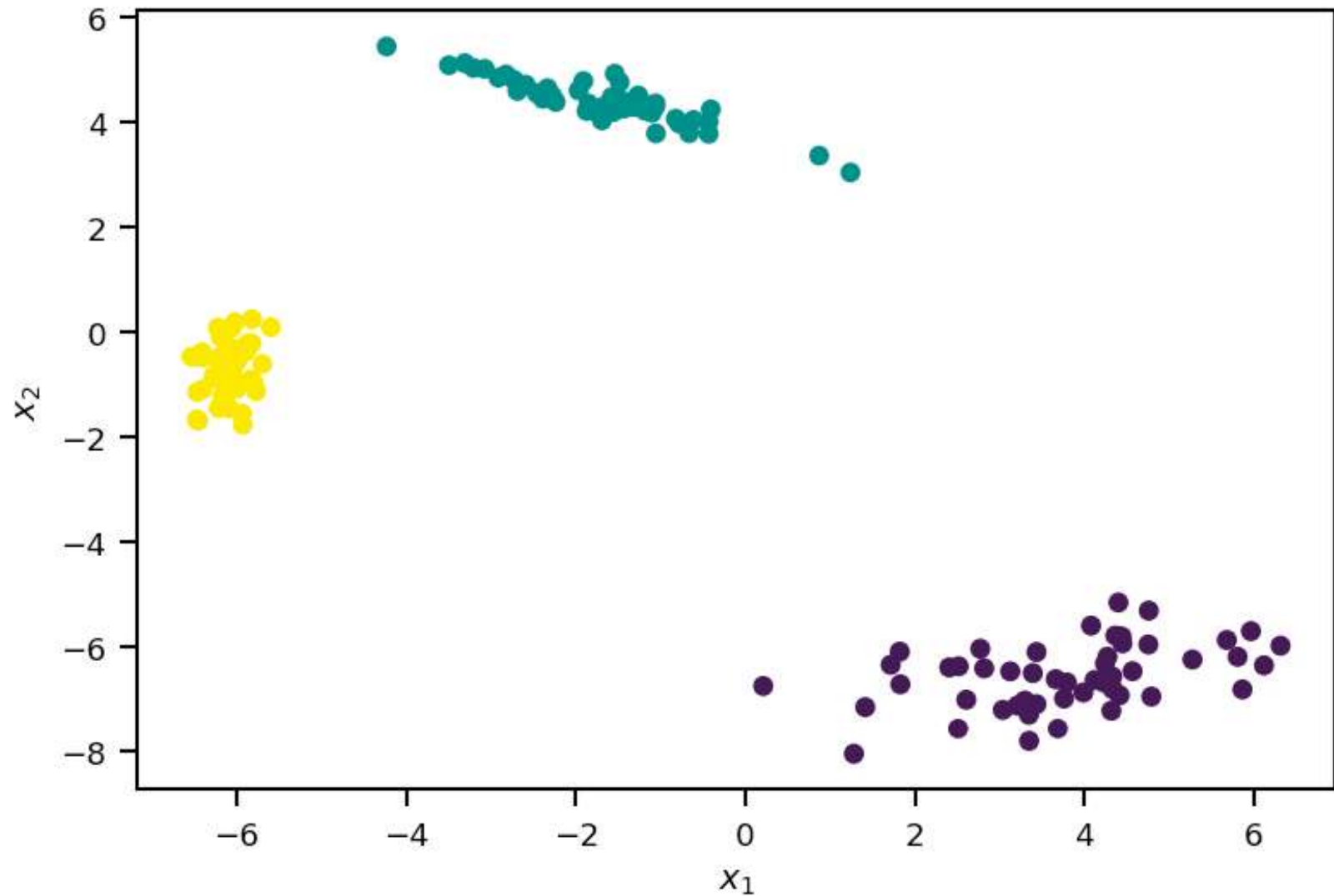


# Lecture 17: Clustering and density estimation

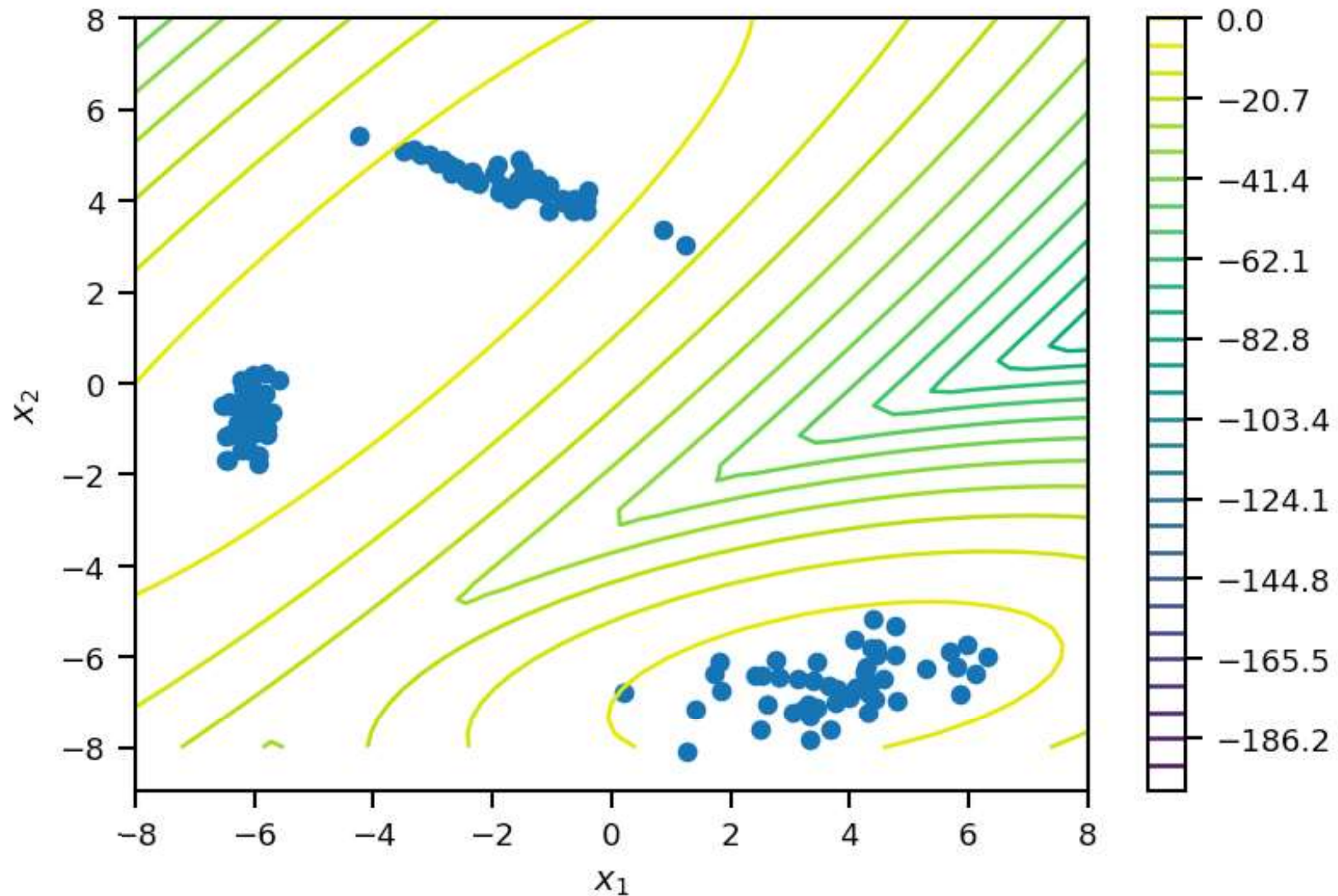
Professor Ilias Bilonis

## Selecting the number of components using the Bayesian information criterion

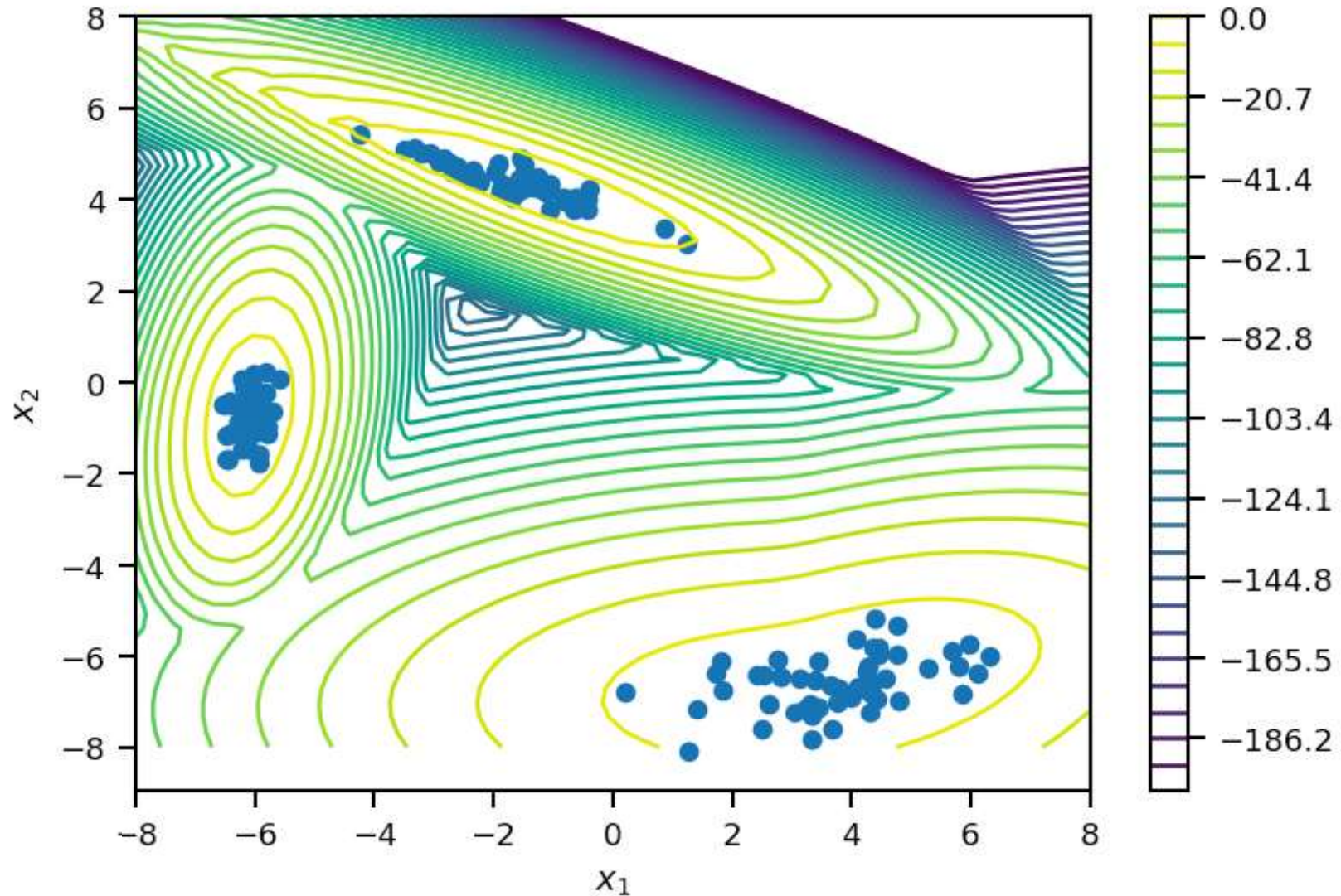
# Example: Clustering with mixtures



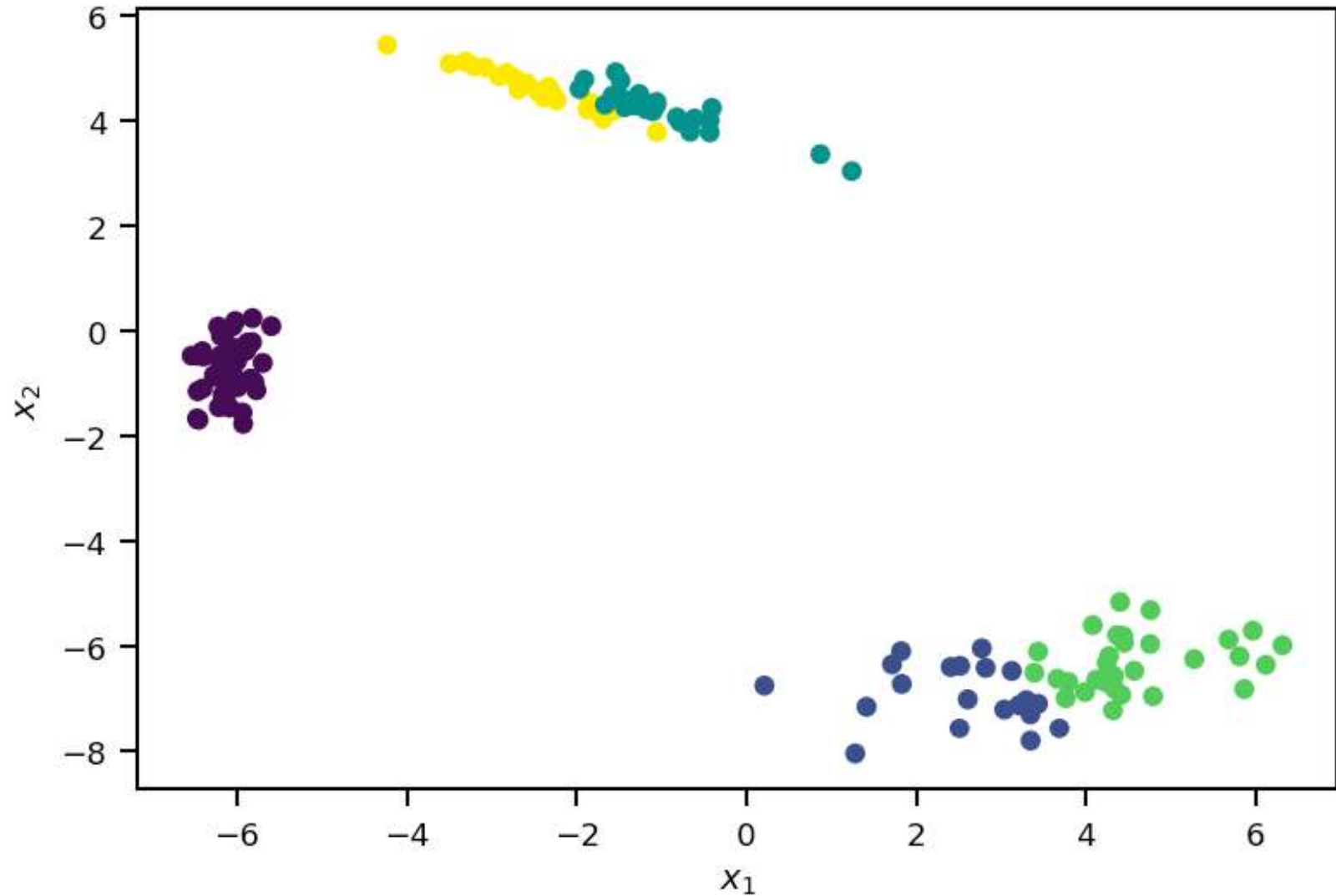
# Trying fewer components doesn't work well



# Trying more components works for density



# But not for clustering





# The Bayesian information criterion

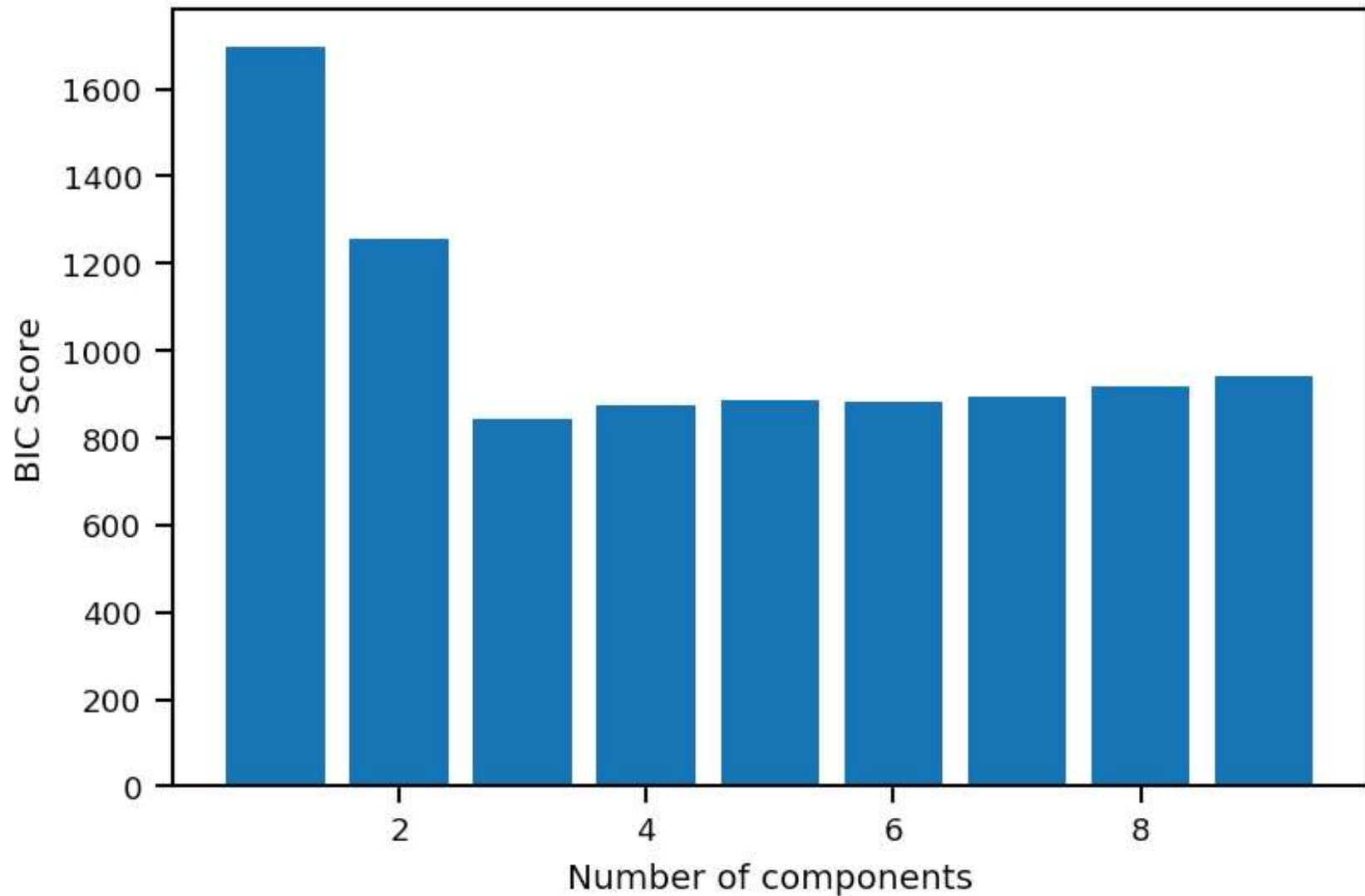
$$BIC = \frac{\# \text{ of model parameters} \cdot \ln n}{\text{Complexity}} - \frac{2 \log\text{-likelihood}}{\text{goodness - A fit. of}}$$

(Bayesian Model Section)  
Lecture 28

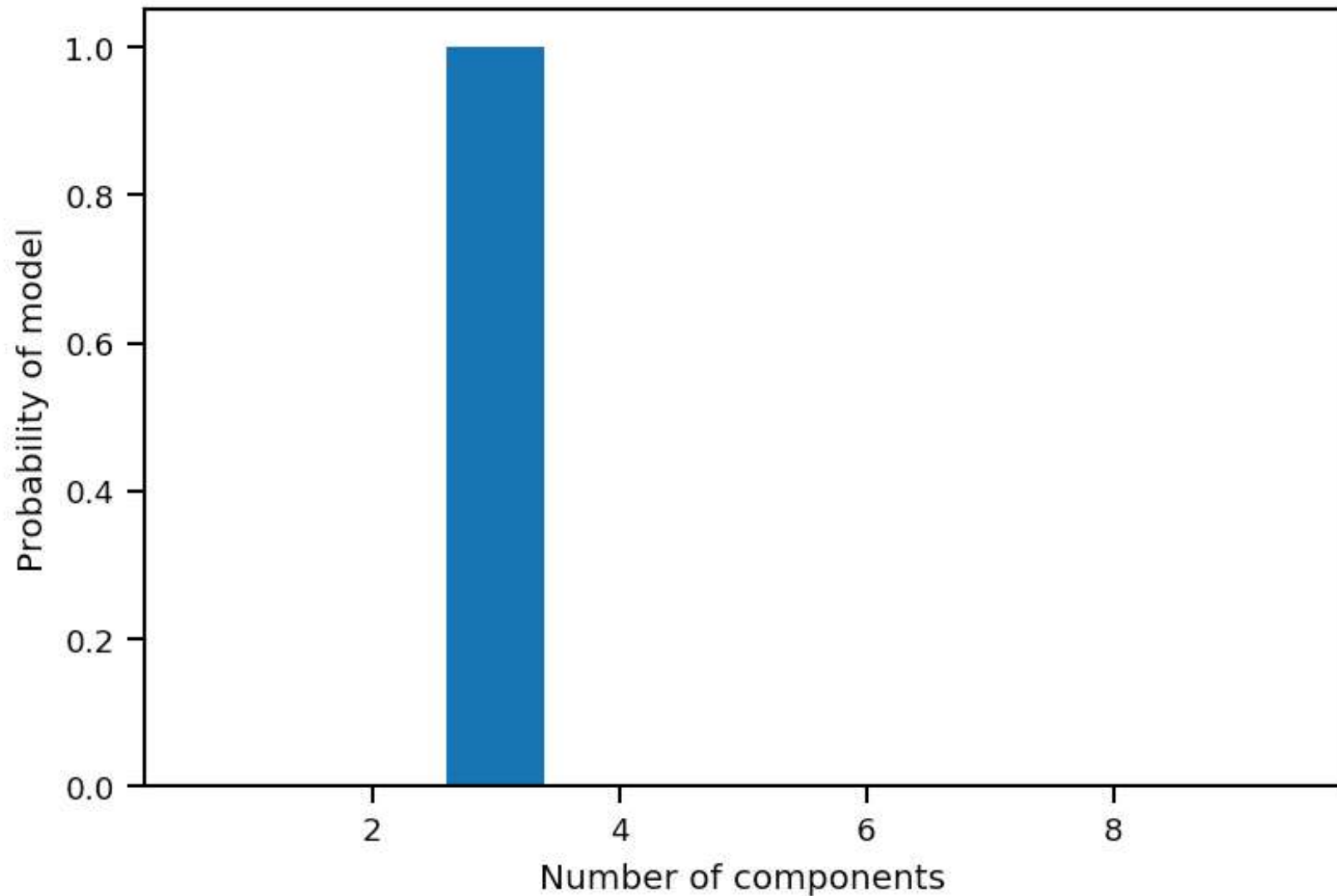
$$K \text{ clusters : } \# \text{ of model parameters} = \underbrace{K-1}_{(I)} + \underbrace{K \cdot D}_{\text{dimension } \neq} + K \cdot D^2$$

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

# BIC scores on our original example

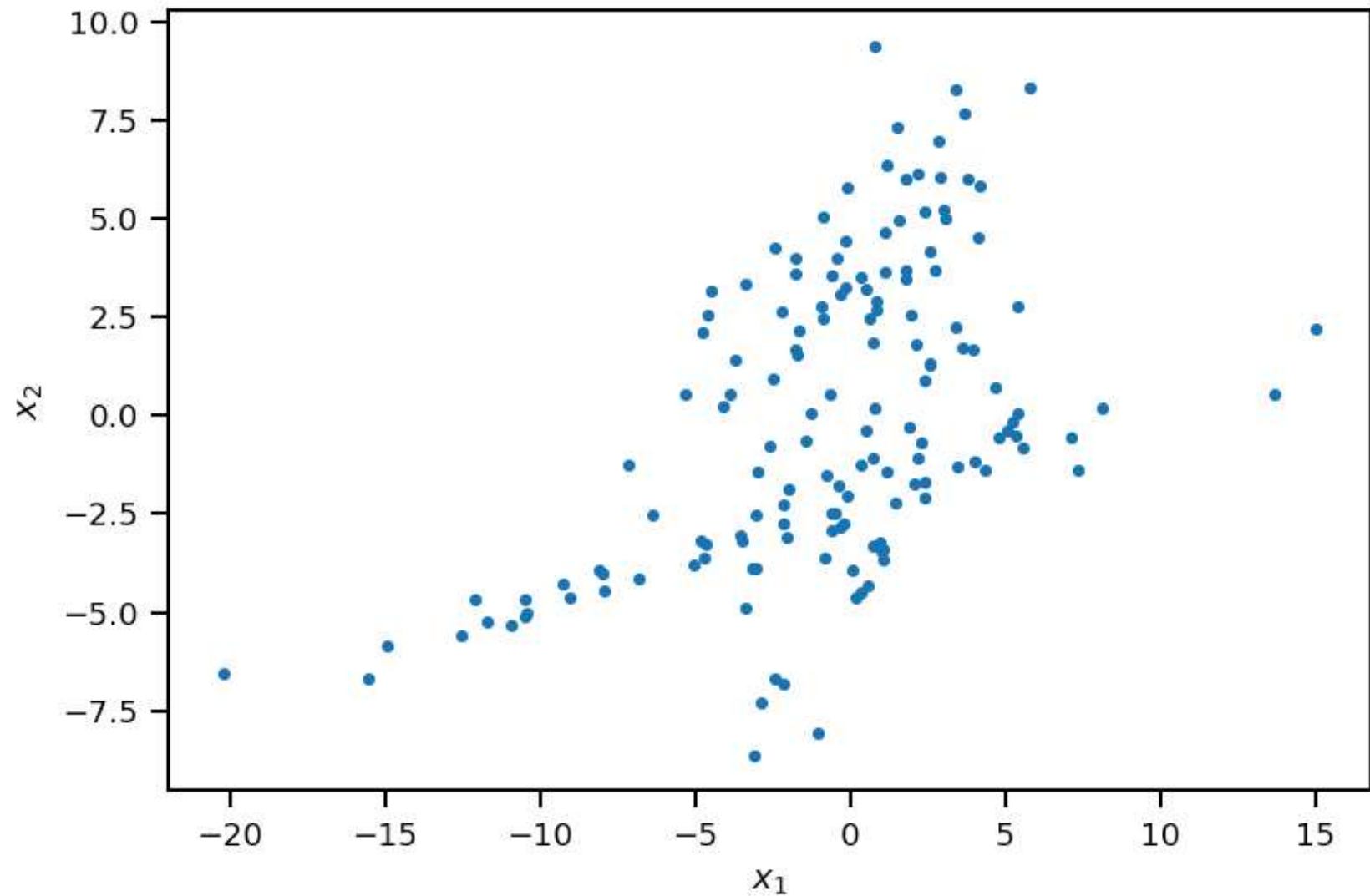


# Probability over number of components

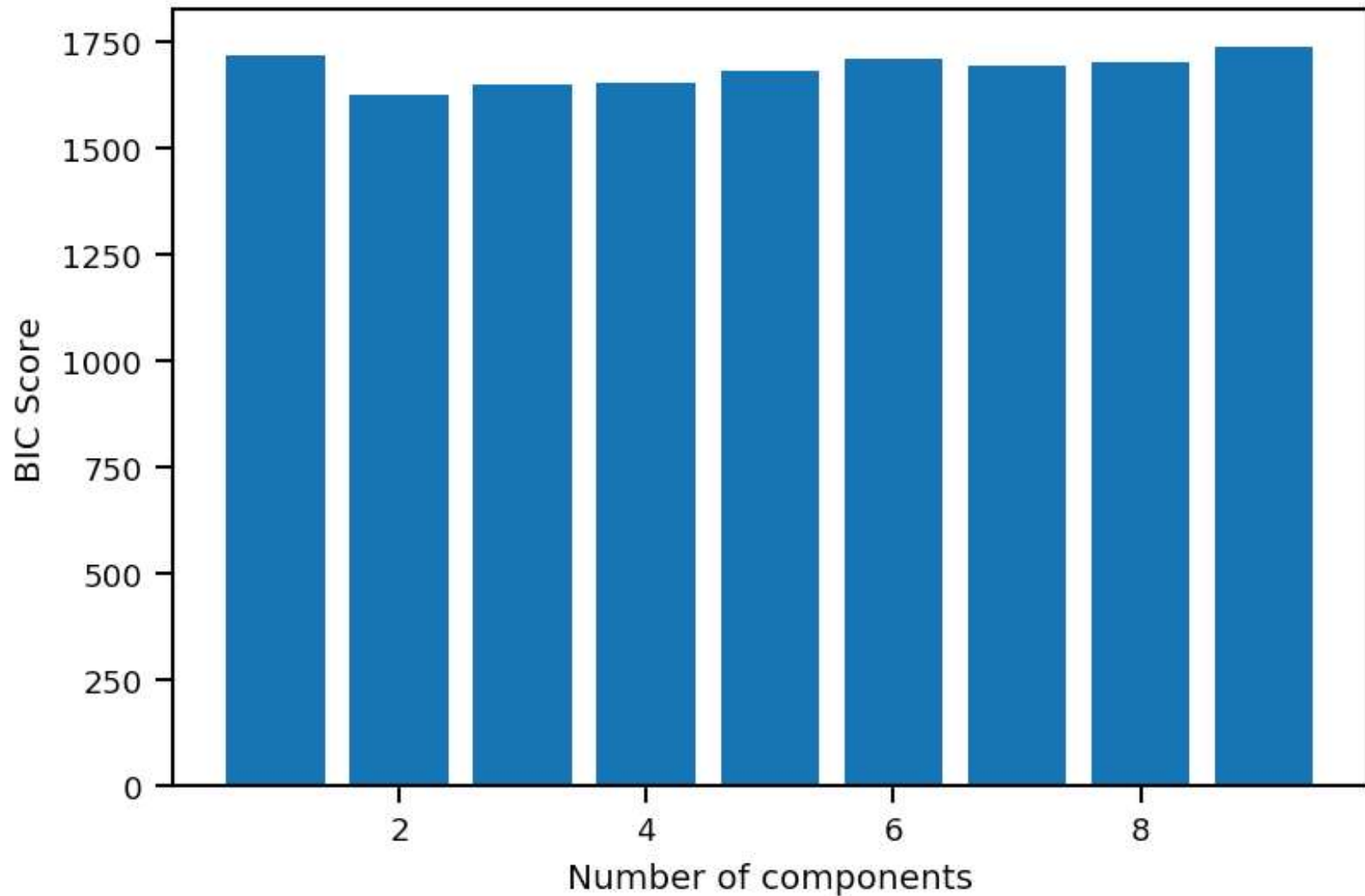




# A less obvious example



# A less obvious example



# Most probable model

