JOINT INSTITUTE
交大密西根学院

VE414 • Bayesian Analysis

Jiuling in the Forbidden Forest

Team Members: Yanjun Zhou, Yangzhen Zhang, Yihao Liu
Instructor: Prof. Jing Liu

Problem Statement

Jiuling trees, a kind of invisible plant, have been in the Forbidden Forest of Hogwarts for many years. With the help of a spell, number of the fruits of Jiuling, Tayes, are recorded along with the positions in some trips. This project is to visualize and analyze the data recorded by the spell and find the number and distribution of Jiuling in the Forbidden Forest.

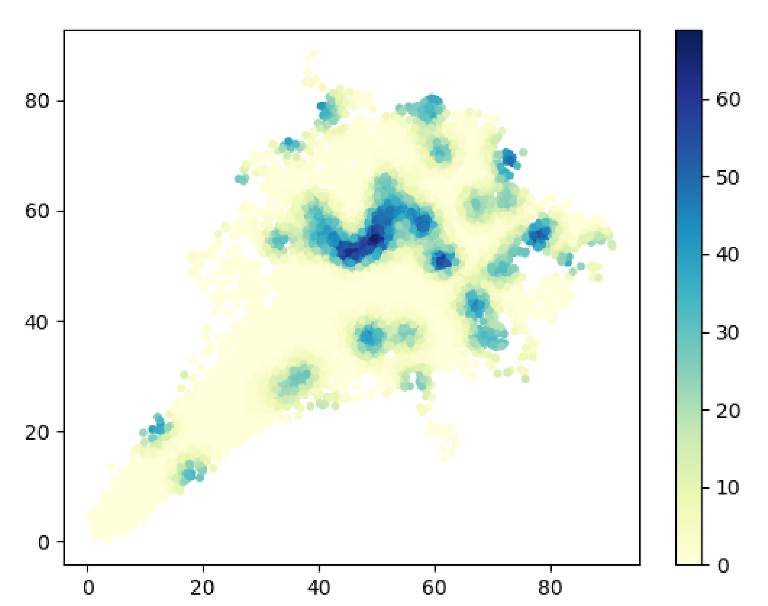


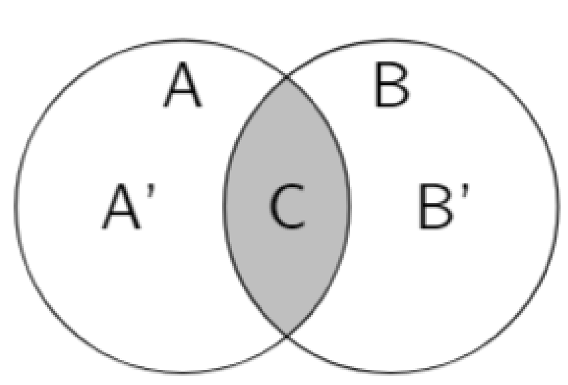
Fig. 1: A simple visualization of the data

Density Estimation

Many records of Tayes $\Longrightarrow$ Estimation of Tayes in the forest

- Divide map into grid

- Calculate the density on the grid



$$\text{Density}[C] \approx \left(\frac{a}{S_A} + \frac{b}{S_B}\right)/2$$

- Sampling

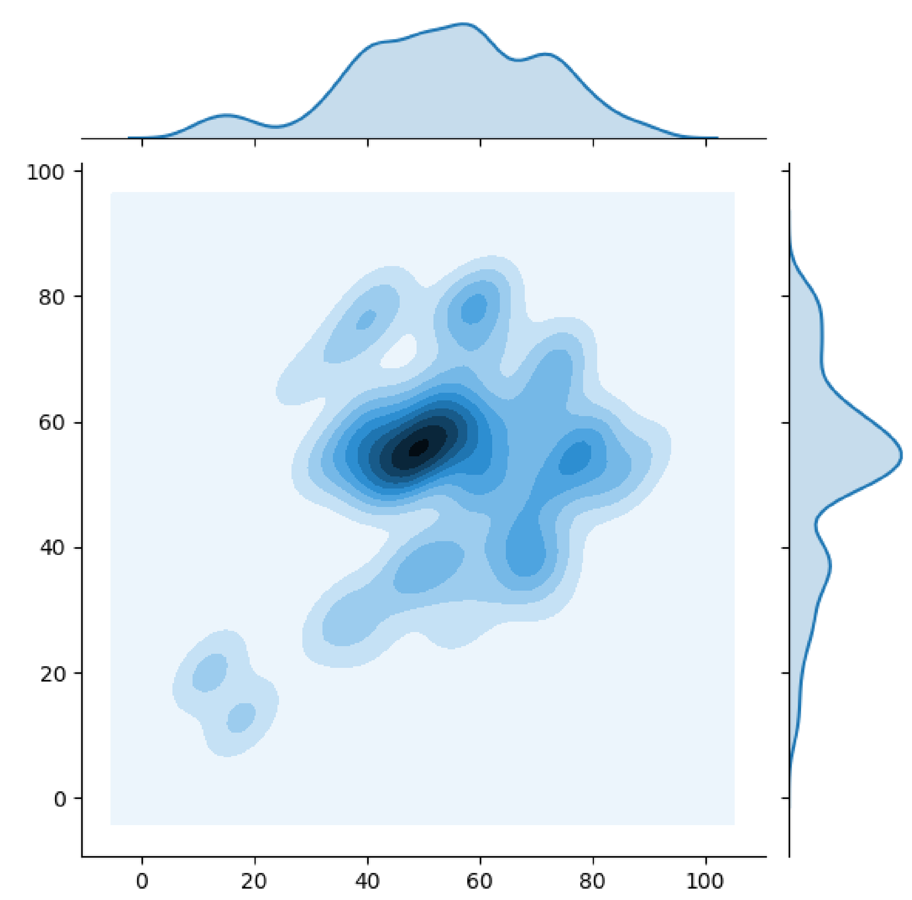


Fig. 2: The density of Tayes over the forest with records

GMM with EM

Apply Expectation-Maximization algorithm (EM) to fit Guassian Mixture Model(GMM) with Tayes Samples.

- E-step: calculate the posterior probabilities for $k \in \{1, \cdots, K\}, n \in \{1, \cdots, N\}$

$$\gamma_{nk} = p(k|x_n) = \frac{\pi_k \mathcal{N}(x_n|\mu_n, \Sigma_n)}{\Sigma_{j=1}^K \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)}$$

- M-step: calculate the value for $\pi_k,\ \mu_k,\ \Sigma_k$

$$\mu_k' = \frac{\Sigma_{n=1}^N \gamma_{nk} x_n}{\Sigma_{n=1}^N \gamma_{nk}}$$

$$\Sigma_k' = \frac{\Sigma_{n=1}^N \gamma_{nk}(x_n - \mu_k')(x_n - \mu_k')^T}{\Sigma_{n=1}^N \gamma_{nk}}$$

$$\pi_k' = \frac{\Sigma_{n=1}^N \gamma_{nk}}{N}$$

- Loop until log likelihood convergence:

$$\ln p(X|\mu, \Sigma, \pi) = \sum_{n=1}^{N} ln\{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)\}$$

- Pick a k = $\arg_k \min$ BIC(Bayesian Information Creterion):

$$BIC = -2ln(L) + ln(N) * K$$

By fitting the density with the Gaussian Mixture Model (GMM), we determined the number of the trees K as 72 based on Bayesian Information Criterion. The mean μ indicates the location of each tree.

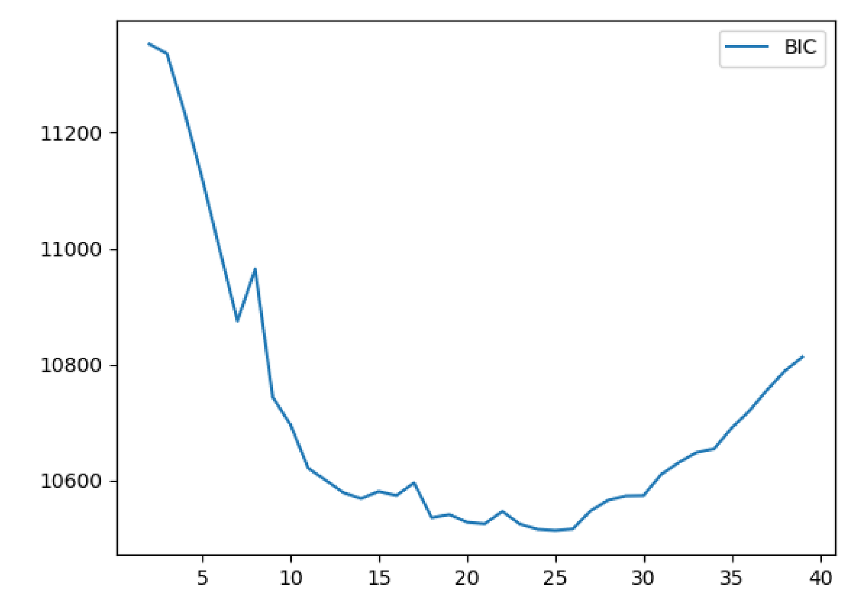


Fig. 3: BIC changes with K

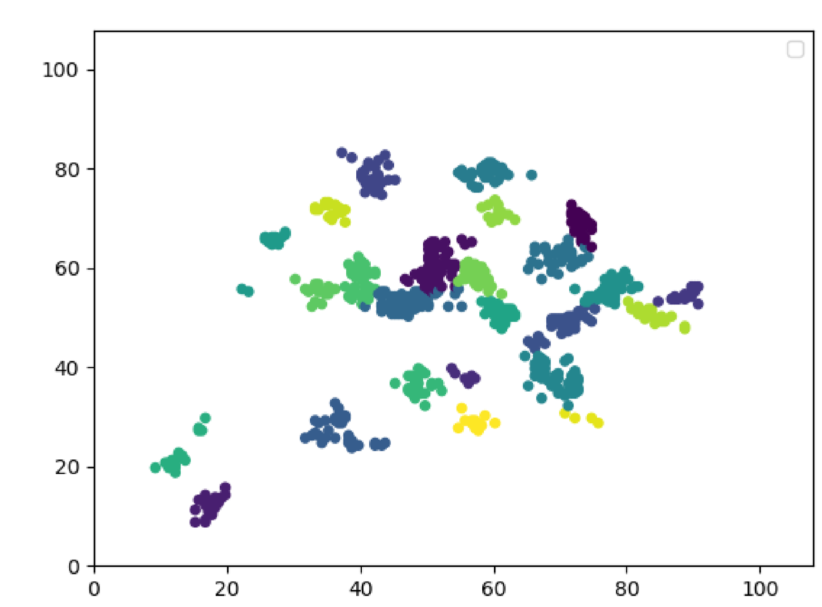


Fig. 4: An example for the tree distribution in the middle of the forest

Prediction

After we obtain a GMM for observed areas, we divide the forest into small grids and fit a model of tree density in each related to its distances to two nearest borders of the forest. Therefore, we can predict the tree density of unobserved areas and make the final prediction of tree distribution and total tree number.

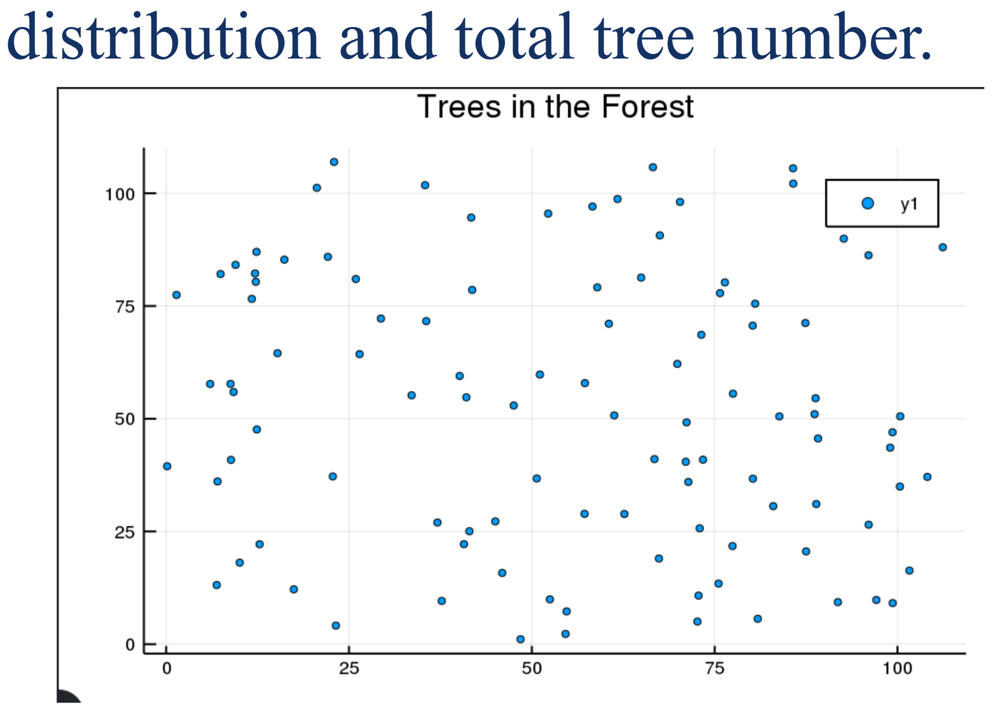


Fig. 5: Prediction of the trees in the forest