Breast Cancer Image Classifications

Yuxuan Chen | Yuan Meng | Paula Wu

Introduction & Data Preprocessing

Breast cancer is only of the most common invasive cancer in women in the Unied States, second only to skin cancer [1]. Starting from different parts of the breast, breast cancer is usually marked by lumpiness or swells in the breasts and surrounding tissues. However, it is important to distinguish benign breast tumors from malignant ones, as non-cancer breast tumors are usually abnormal growths that do not spread outside of the breast. They are not life-threatening, even though some types of benign breast lumps can increase a women's risk of getting breast cancer [2]. On the other hand, malignant tumors are aggressive and deadly. Fortunately, the prognosis of the disease has been greatly improved once it is detected. Therefore, it's important to have diseased tissue accurately diagnosed, as misdiagnosis may leads to delayed intervention or may cause unnecessary stress on a patient.

Data Overview

The original dataset contains 569 observations and 33 columns [3]. Among the 569 observations, there are 212 malignant observations and 357 benign observations. Thirty out of the 33 columns are predictors regarding imaging features, computed from a digitized image of a fine needle aspirate (FNA) of a breast mass [4]. They describe the characteristics of the cell nuclei present in the image. Ten real-valued features are computed for each cell nucleus: radius (mean of distances from the center to points on the perimeter), texture (standard deviation of gray-scale values), perimeter, area, smoothness (local variation in radius lengths), compactness (perimeter^2 / area - 1.0), concavity (severity of concave portions of the contour), concave points (number of concave portions of the contour), symmetry, and fractal dimension ("coastline approximation" - 1). The mean, standard error, and "worst" or largest (mean of the three largest values) of these features that were computed for each image are included in this dataset. In this project, our goal is trying to determine which features contribute to a better diagnosis of breast tissues (M = malignant, B = benign).

Exploratory Data Analysis

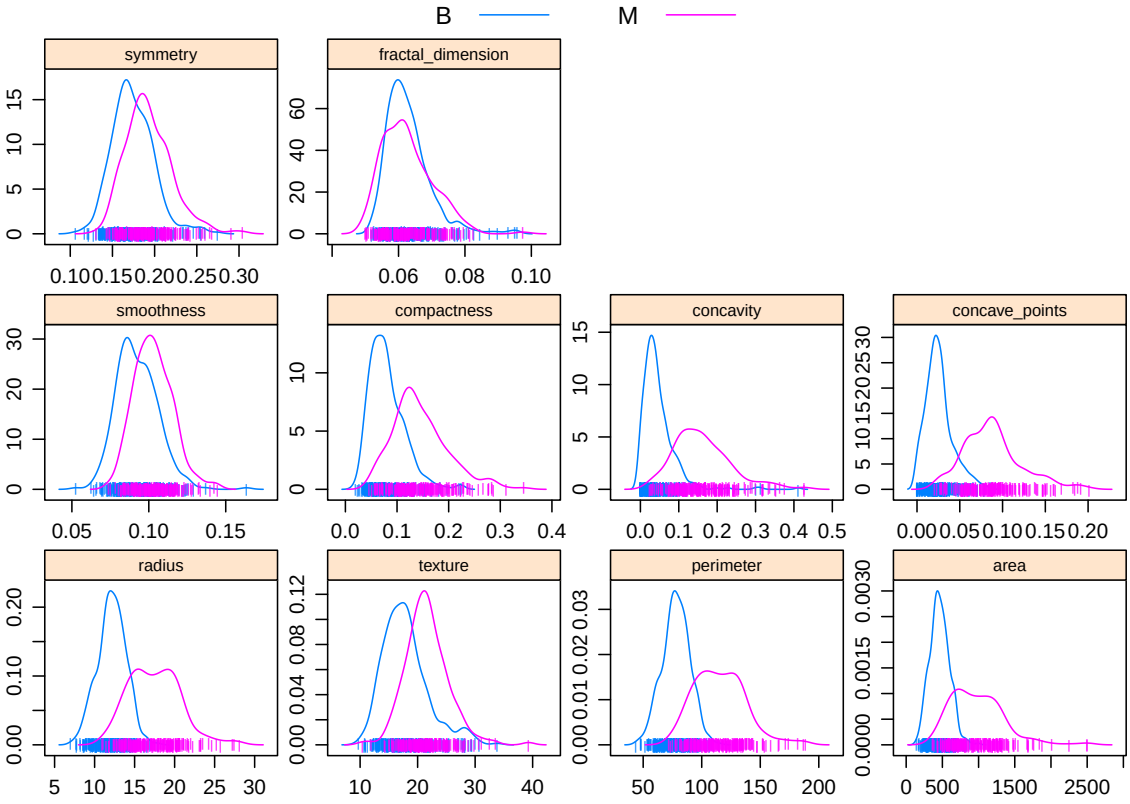
Modeling & Results

aha section

Conclusion & Discussion

Appendix

```
## [1] M B
## Levels: B M
```



Hyper-parameter Tuning for Selected Models

References

[1] Breast cancer basic information.

[2] Lecture Note 7 p2-3

[3] 3

[4] 4

[5] 5