CS7641 Fall 2018 Assignment1 Supervised learning

Xiaoxi Wang

Xwang738

Datasets

Size Instances

Splitting into training (70%) and testing (30%)

5 folds cross validation for hyperparameter tuning

Decision Tree

Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image.

k-Nearest Neighbors

Model Complexity curve over k

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Task | #Instances | #Attributes |  |  |  |  |  |  |
| HTRU2 | Binary Classification |  |  |  |  |  |  |  |  |
| Breast Cancer Wisconsin | Binary Classification | 683\* | 9\*\* | Real |  |  |  |  |  |

\*Removed 16 instances with missing data  
\*\*ID number attribute was dropped

Reference

W.N. Street, W.H. Wolberg and O.L. Mangasarian. Nuclear feature extraction for breast tumor diagnosis. IS&T/SPIE 1993 International Symposium on Electronic Imaging: Science and Technology, volume 1905, pages 861-870, San Jose, CA, 1993.   
[[Web Link]](http://rexa.info/paper/b98475235164960529ad2ff9fda3816e9335cf8a)   
  
O.L. Mangasarian, W.N. Street and W.H. Wolberg. Breast cancer diagnosis and prognosis via linear programming. Operations Research, 43(4), pages 570-577, July-August 1995.   
[[Web Link]](http://rexa.info/paper/90e988e83c7f06d2797b41580569c1f9a13f6749)