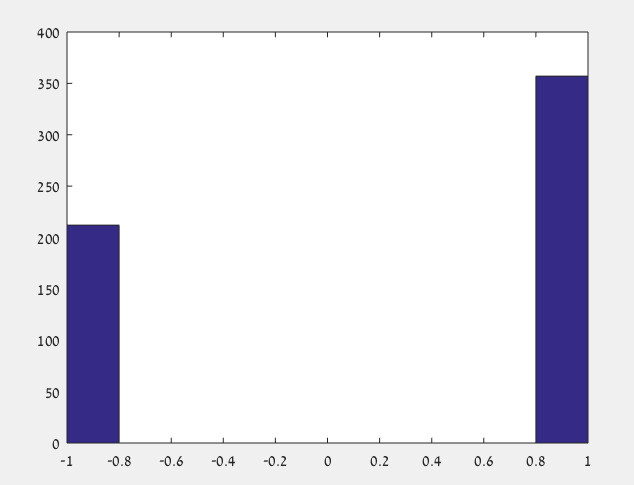
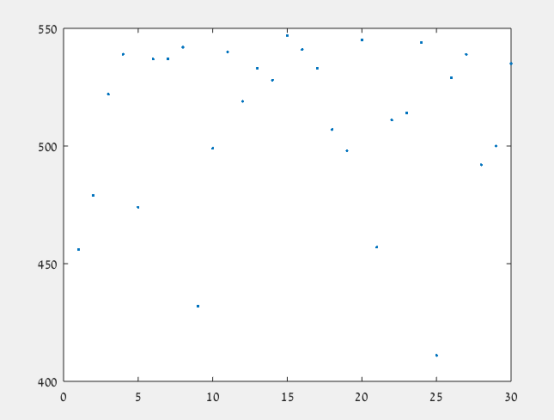
Breast Cancer Wisconsin

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 569 | **Area:** | Life |
| **Attribute Characteristics:** | Real | **Number of Attributes:** | 30 | **Date Donated** | 1995-11-01 |
| **Associated Tasks:** | Classification | **Missing Values?** | No | **Number of Web Hits:** | 348825 |

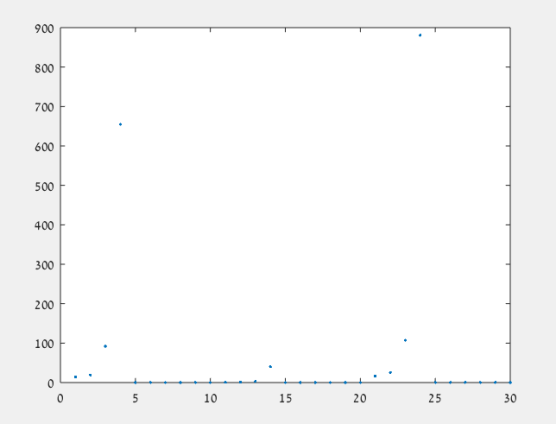
Besov smoothness 



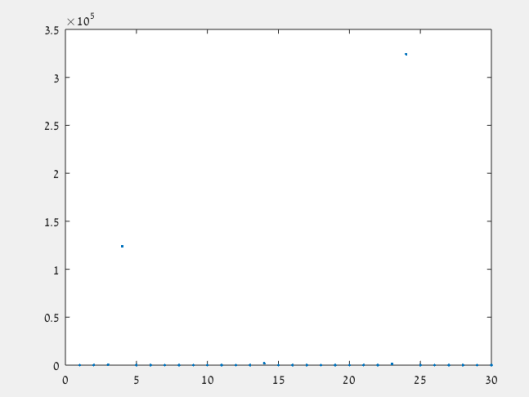
Binary data {-1,1} with mean 0.2548



Unique data (per variable)



Mean data (per variable)



Variance (per variable)

**Data Set Information:**

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. A few of the images can be found at [[Web Link]](http://www.cs.wisc.edu/%7Estreet/images/)   
  
Separating plane described above was obtained using Multisurface Method-Tree (MSM-T) [K. P. Bennett, "Decision Tree Construction Via Linear Programming." Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society, pp. 97-101, 1992], a classification method which uses linear programming to construct a decision tree. Relevant features were selected using an exhaustive search in the space of 1-4 features and 1-3 separating planes.   
  
The actual linear program used to obtain the separating plane in the 3-dimensional space is that described in: [K. P. Bennett and O. L. Mangasarian: "Robust Linear Programming Discrimination of Two Linearly Inseparable Sets", Optimization Methods and Software 1, 1992, 23-34].   
  
This database is also available through the UW CS ftp server:   
ftp ftp.cs.wisc.edu   
cd math-prog/cpo-dataset/machine-learn/WDBC/

**Attribute Information:**

1) ID number   
2) Diagnosis (M = malignant, B = benign)   
3-32)   
  
Ten real-valued features are computed for each cell nucleus:   
  
a) radius (mean of distances from center to points on the perimeter)   
b) texture (standard deviation of gray-scale values)   
c) perimeter   
d) area   
e) smoothness (local variation in radius lengths)   
f) compactness (perimeter^2 / area - 1.0)   
g) concavity (severity of concave portions of the contour)   
h) concave points (number of concave portions of the contour)   
i) symmetry   
j) fractal dimension ("coastline approximation" - 1)

**Relevant Papers:**

First Usage:   
  
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)   
  
Medical literature:   
  
W.H. Wolberg, W.N. Street, and O.L. Mangasarian. Machine learning techniques to diagnose breast cancer from fine-needle aspirates. Cancer Letters 77 (1994) 163-171.   
[[Web Link]](http://rexa.info/paper/6b740b6a2f0cd6dc78ef18d9044dc51990b4143e)   
  
W.H. Wolberg, W.N. Street, and O.L. Mangasarian. Image analysis and machine learning applied to breast cancer diagnosis and prognosis. Analytical and Quantitative Cytology and Histology, Vol. 17 No. 2, pages 77-87, April 1995.   
  
W.H. Wolberg, W.N. Street, D.M. Heisey, and O.L. Mangasarian. Computerized breast cancer diagnosis and prognosis from fine needle aspirates. Archives of Surgery 1995;130:511-516.   
[[Web Link]](http://rexa.info/paper/14d31bfd5fbf5b1ff093832f1594cb0418a3a4ec)   
  
W.H. Wolberg, W.N. Street, D.M. Heisey, and O.L. Mangasarian. Computer-derived nuclear features distinguish malignant from benign breast cytology. Human Pathology, 26:792--796, 1995.   
[[Web Link]](http://rexa.info/paper/8c8d827f59d2982b4865f5a47e9607df3a7489a4)