Cancer Dataset
Wisconsin Diagnostic Breast Cancer (WDBC)

Source: Creators:

Dr. William H. Wolberg, General Surgery Dept., University of Wisconsin
W. Nick Street, Computer Sciences Dept., University of Wisconsin
Olvi L. Mangasarian, Computer Sciences Dept., University of Wisconsin

Date: November 1995

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.

Objective: predicting diagnosis field with two values: B = benign, M = malignant

Relevant 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 http://www.cs.wisc.edu/~street/images/

Number of instances: 569

Number of attributes: 32 (ID, diagnosis, 30 real-valued input features)

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)
Then we have the same set of values but their standard error (RadiusSE) and worst (RadiusMax, (mean of the three largest values) values.

Missing attribute values: none

Class distribution: 357 benign, 212 malignant