

- Home
- Compete
- Data
- Notebooks
- Discuss
- Courses
- More

- Recently Viewed
- Breast Cancer Wiscon...
  - COVID-19 Literature Cl...
  - COVID-19 Case Study ...
  - Building an AI-powere...
  - Questions about Table ...

DataTasks (2)Kernels (1,113)Discussion (27)ActivityMetadata

Download (122 KB)New Notebook

Dataset

1461

### Breast Cancer Wisconsin (Diagnostic) Data Set

Predict whether the cancer is benign or malignant

UCI Machine Learning • updated 4 years ago (Version 2)

Your Dataset download has started.  
Show your appreciation with an upvote

1461

Usability 8.5License CC BY-NC-SA 4.0Tags mathematics, cancer, healthcare

Description

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. n 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/

Also can be found on UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>

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)

The mean, standard error and "worst" or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. For instance, field 3 is Mean Radius, field 13 is Radius SE, field 23 is Worst Radius.

All feature values are recoded with four significant digits.

Missing attribute values: none

Class distribution: 357 benign, 212 malignant

Data Explorer

122.27 KB

data.csv

Summary

1 file

32 columns

< data.csv (122.27 KB)

DetailCompactColumn

10 of 32 columns

About this file

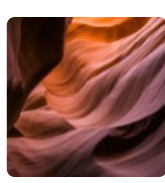
This file contains all the data in

id	diagnosis	radius_mean	texture_mean	perimeter_mean	
ID number	The diagnosis of breast tissues (M = malignant, B = benign)	mean of distances from center to points on the perimeter	standard deviation of gray-scale values	mean size of the core tumor	
	B M	63% 37%			
8670		6.98	9.71	43.8	104
842302	M	17.99	10.38	122.8	106
842517	M	28.57	17.77	132.9	132
84380903	M	19.69	21.25	138	126
84348301	M	11.42	20.38	77.58	386
84358402	M	28.29	14.34	135.1	125
843786	M	12.45	15.7	82.57	477
844359	M	18.25	19.98	119.6	184
84458202	M	13.71	28.83	98.2	577
844981	M	13	21.82	87.5	515
84581001	M	12.46	24.84	83.97	475
845636	M	16.82	23.24	182.7	797
84610002	M	15.78	17.89	183.6	781
846226	M	19.17	24.8	132.4	112
846381	M	15.85	23.95	183.7	782
84667401	M	13.73	22.61	93.6	576
84799002	M	14.54	27.54	96.73	656
848406	M	14.68	28.13	94.74	684
84862001	M	16.13	28.68	188.1	796
849014	M	19.81	22.15	138	126
8510426	B	13.54	14.36	87.46	566
8510653	B	13.88	15.71	85.63	526
8510824	B	9.584	12.44	68.34	273
8511133	M	15.34	14.26	182.5	784
851509	M	21.16	23.84	137.2	146
852552	M	16.65	21.38	118	984
852631	M	17.14	16.4	116	912
852763	M	14.58	21.53	97.41	644
852781	M	18.61	28.25	122.1	106
852973	M	15.3	25.27	182.4	732
853201	M	17.57	15.85	115	956
853401	M	18.63	25.11	124.8	106
853612	M	11.84	18.7	77.93	446
85382601	M	17.82	23.98	112.8	895
854002	M	19.27	26.47	127.9	116
854039	M	16.13	17.88	187	887
854253	M	16.74	21.59	118.1	865
854268	M	14.25	21.72	93.63	632
854941	B	13.83	18.42	82.61	522
855133	M	14.99	25.2	95.54	696
855138	M	13.48	28.82	88.4	555
855167	M	13.44	21.58	86.18	565
855563	M	18.95	21.35	71.9	371
855625	M	19.87	24.81	128.3	116
856106	M	13.28	28.28	87.32	545
85638502	M	13.17	21.81	85.42	531
857010	M	18.65	17.6	123.7	187
85713702	B	8.196	16.84	51.71	281
85715	M	13.17	18.66	85.98	534
857155	B	12.85	14.63	78.84	445
857156	B	13.49	22.3	86.91	561

Similar Datasets



Google Play Store Apps



Pima Indians Diabetes Database



Brazilian E-Commerce Public Dataset By Olist



Malaria Cell Images Dataset



Medical Cost Personal Datasets

558,956 views · 95,007 downloads · 1,113 kernels · 27 topics

[View more activity](#)