The data set (and description) can be downloaded here:
http://www.stats.ox.ac.uk/pub/PRNN/pima.tr

Description:

```
1. Title: Pima Indians Diabetes Database

2. Sources:
   (a) Original owners: National Institute of Diabetes and Digestive and
                        Kidney Diseases
   (b) Donor of database: Vincent Sigillito (vgs@aplcen.apl.jhu.edu)
                          Research Center, RMI Group Leader
                          Applied Physics Laboratory
                          The Johns Hopkins University
                          Johns Hopkins Road
                          Laurel, MD 20707
                          (301) 953-6231
   (c) Date received: 9 May 1990

3. Past Usage:
    1. Smith,~J.~W., Everhart,~J.~E., Dickson,~W.~C., Knowler,~W.~C., \&
       Johannes,~R.~S. (1988). Using the ADAP learning algorithm to forecast
       the onset of diabetes mellitus.  In {\it Proceedings of the Symposium
       on Computer Applications and Medical Care} (pp. 261--265).  IEEE
       Computer Society Press.

       The diagnostic, binary-valued variable investigated is whether the
       patient shows signs of diabetes according to World Health Organization
       criteria (i.e., if the 2 hour post-load plasma glucose was at least
       200 mg/dl at any survey  examination or if found during routine medical
       care).   The population lives near Phoenix, Arizona, USA.

       Results: Their ADAP algorithm makes a real-valued prediction between
       0 and 1.  This was transformed into a binary decision using a cutoff of
       0.448.  Using 576 training instances, the sensitivity and specificity
       of their algorithm was 76% on the remaining 192 instances.

4. Relevant Information:
      Several constraints were placed on the selection of these instances from
      a larger database.  In particular, all patients here are females at
      least 21 years old of Pima Indian heritage.  ADAP is an adaptive learning
      routine that generates and executes digital analogs of perceptron-like
      devices.  It is a unique algorithm; see the paper for details.

5. Number of Instances: 768

6. Number of Attributes: 8 plus class

7. For Each Attribute: (all numeric-valued)
   1. Number of times pregnant
   2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test
   3. Diastolic blood pressure (mm Hg)
   4. Triceps skin fold thickness (mm)
   5. 2-Hour serum insulin (mu U/ml)
   6. Body mass index (weight in kg/(height in m)^2)
   7. Diabetes pedigree function
   8. Age (years)
   9. Class variable (0 or 1)

8. Missing Attribute Values: Yes
```

9. Class Distribution: (class value 1 is interpreted as "tested positive for diabetes")

Class Value	Number of instances
0	500
1	268

10. Brief statistical analysis:

Attribute number:	Mean:	Standard Deviation:
1.	3.8	3.4
2.	120.9	32.0
3.	69.1	19.4
4.	20.5	16.0
5.	79.8	115.2
6.	32.0	7.9
7.	0.5	0.3
8.	33.2	11.8

Descriptive statistics:

Dataset= pima : n= 200 , d= 7

Class1: n= 132

Covariance matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	7.8785	10.7723	8.1908	2.9103	-0.0358	-0.2073	16.8289
[2,]	10.7723	709.5612	81.4303	13.2377	19.0379	-0.5186	59.0131
[3,]	8.1908	81.4303	122.8452	33.8799	16.6126	-0.0772	46.7870
[4,]	2.9103	13.2377	33.8799	119.4464	50.1259	0.0743	18.4707
[5,]	-0.0358	19.0379	16.6126	50.1259	40.7230	0.1452	7.0000
[6,]	-0.2073	-0.5186	-0.0772	0.0743	0.1452	0.0714	-0.5381
[7,]	16.8289	59.0131	46.7870	18.4707	7.0000	-0.5381	91.0818

Correlation matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	1.0000	0.1441	0.2633	0.0949	-0.0020	-0.2765	0.6282
[2,]	0.1441	1.0000	0.2758	0.0455	0.1120	-0.0729	0.2321
[3,]	0.2633	0.2758	1.0000	0.2797	0.2349	-0.0261	0.4423
[4,]	0.0949	0.0455	0.2797	1.0000	0.7187	0.0254	0.1771
[5,]	-0.0020	0.1120	0.2349	0.7187	1.0000	0.0852	0.1149
[6,]	-0.2765	-0.0729	-0.0261	0.0254	0.0852	1.0000	-0.2110
[7,]	0.6282	0.2321	0.4423	0.1771	0.1149	-0.2110	1.0000

Median:	2.5968	110.4798	68.5842	26.3021	30.9338	0.4076	27.3908
Mean:	2.9167	113.1061	69.5455	27.2045	31.0742	0.4155	29.2348
MCD-estimated:							
MDC-0.975-Mean:	1.9579	107.7895	67.4	25.3474	30.6074	0.4068	24.7263
MDC-0.750-Mean:	1.9579	107.7895	67.4	25.3474	30.6074	0.4068	24.7263
MDC-0.500-Mean:	1.8602	107.6882	67.3441	25.2903	30.6419	0.4078	24.6667

```
Class2: n= 68

Covariance matrix:
        [,1]     [,2]     [,3]     [,4]    [,5]    [,6]     [,7]
[1,] 15.7794  -8.1993   6.4249  -0.5180 -1.0329 -0.1330  21.9344
[2,] -8.1993 907.2502  23.7112  87.5154  9.9816 -0.0822  58.1229
[3,]  6.4249  23.7112 134.1861  19.7059  5.1589 -0.7955  26.3038
[4,] -0.5180  87.5154  19.7059 151.3292 28.2855  0.3480  26.6787
[5,] -1.0329   9.9816   5.1589  28.2855 23.1453  0.4577  -7.9122
[6,] -0.1330  -0.0822  -0.7955   0.3480  0.4577  0.1289  -0.4174
[7,] 21.9344  58.1229  26.3038  26.6787 -7.9122 -0.4174 131.7987

Correlation matrix:
        [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]
[1,]  1.0000 -0.0685  0.1396 -0.0106 -0.0540 -0.0933  0.4810
[2,] -0.0685  1.0000  0.0680  0.2362  0.0689 -0.0076  0.1681
[3,]  0.1396  0.0680  1.0000  0.1383  0.0926 -0.1913  0.1978
[4,] -0.0106  0.2362  0.1383  1.0000  0.4779  0.0788  0.1889
[5,] -0.0540  0.0689  0.0926  0.4779  1.0000  0.2650 -0.1433
[6,] -0.0933 -0.0076 -0.1913  0.0788  0.2650  1.0000 -0.1013
[7,]  0.4810  0.1681  0.1978  0.1889 -0.1433 -0.1013  1.0000

Median:           5.1258 144.4749 74.9284 33.0938 34.7369 0.5428 37.4477

Mean:             4.8382 145.0588 74.5882 33.1176 34.7088 0.5487 37.6912
MCD-estimated:
MDC-0.975-Mean:   5.4737 144.3509 74.386  32.193  34.1404 0.499  38.1053
MDC-0.750-Mean:   5.2778 143.9259 75.1111 32.8704 34.3463 0.4983 37.6111
MDC-0.500-Mean:   5.0847 144.1186 75.4576 32.1695 34.4593 0.4867 37.0169

Measures:
Mah.Dist:                        1.5192
Mah.Dist-MCD-0.975:              2.1257
Mah.Dist-MCD-0.750:              2.2001
Mah.Dist-MCD-0.500:              2.1567
```