The data set (and description) can be downloaded here:
http://archive.ics.uci.edu/ml/machine-learning-databases/pima-indians-diabetes/pima-indians-diabetes.data

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

Citation Request:

Please refer to the repository http://archive.ics.uci.edu/ml (see citation policy).
See also Frank, A. & Asuncion, A. (2010). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml].
Irvine, CA: University of California, School of Information and Computer Science.

Descriptive statistics:

Dataset= diabetes : n= 768 , d= 8

Class1: n= 268

Covariance matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]
[1,]	13.9969	-6.5233	10.2086	-5.2363	-40.7637	-4.3224	-0.0964	18.2600
[2,]	-6.5233	1020.1395	47.1577	21.2422	1157.7746	11.6957	0.3149	34.5295
[3,]	10.2086	47.1577	461.8980	85.5870	266.3737	20.9090	0.2763	61.9219
[4,]	-5.2363	21.2422	85.5870	312.5722	1119.4727	40.0705	1.8031	-17.8425
[5,]	-40.7637	1157.7746	266.3737	1119.4727	19234.6733	55.5141	5.2450	36.4231
[6,]	-4.3224	11.6957	20.9090	40.0705	55.5141	52.7507	0.3699	-14.9774
[7,]	-0.0964	0.3149	0.2763	1.8031	5.2450	0.3699	0.1386	-0.3599
[8,]	18.2600	34.5295	61.9219	-17.8425	36.4231	-14.9774	-0.3599	120.3026

Correlation matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]
[1,]	1.0000	-0.0546	0.1270	-0.0792	-0.0786	-0.1591	-0.0692	0.4450
[2,]	-0.0546	1.0000	0.0687	0.0376	0.2614	0.0504	0.0265	0.0986
[3,]	0.1270	0.0687	1.0000	0.2252	0.0894	0.1340	0.0345	0.2627
[4,]	-0.0792	0.0376	0.2252	1.0000	0.4566	0.3121	0.2739	-0.0920
[5,]	-0.0786	0.2614	0.0894	0.4566	1.0000	0.0551	0.1016	0.0239
[6,]	-0.1591	0.0504	0.1340	0.3121	0.0551	1.0000	0.1368	-0.1880
[7,]	-0.0692	0.0265	0.0345	0.2739	0.1016	0.1368	1.0000	-0.0881
[8,]	0.4450	0.0986	0.2627	-0.0920	0.0239	-0.1880	-0.0881	1.0000

Median:	4.8835	137.5659	71.6751	20.1639	59.7598	34.7332	0.5374	36.6408
Mean:	4.8657	141.2575	70.8246	22.1642	100.3358	35.1425	0.5505	37.0672
MCD-estimated:								
MDC-0.975-Mean:	5.2391	137.5507	67.7609	11.9928	0	34.5442	0.4798	38.1304
MDC-0.750-Mean:	5.2391	137.5507	67.7609	11.9928	0	34.5442	0.4798	38.1304
MDC-0.500-Mean:	5.2391	137.5507	67.7609	11.9928	0	34.5442	0.4798	38.1304

```
Class2: n= 500

Covariance matrix:
         [,1]     [,2]     [,3]     [,4]      [,5]     [,6]    [,7]      [,8]
[1,]   9.1034   7.7835   7.2537  -5.3165  -39.3708   0.3827 -0.0721   20.1637
[2,]   7.7835 683.3623  91.0358   6.2337  912.2022  26.4845  0.7470   69.5469
[3,]   7.2537  91.0358 326.2747  50.3145  133.2688  50.4463  0.1474   45.2475
[4,]  -5.3165   6.2337  50.3145 221.7105  607.6674  50.2211  0.4239  -28.4551
[5,] -39.3708 912.2022 133.2688 607.6674 9774.3454 193.2592  6.7236 -172.1448
[6,]   0.3827  26.4845  50.4463  50.2211  193.2592  59.1339  0.1625    3.2363
[7,]  -0.0721   0.7470   0.1474   0.4239    6.7236   0.1625  0.0895    0.1454
[8,]  20.1637  69.5469  45.2475 -28.4551 -172.1448   3.2363  0.1454  136.1342

Correlation matrix:
         [,1]   [,2]   [,3]    [,4]    [,5]   [,6]    [,7]    [,8]
[1,]  1.0000 0.0987 0.1331 -0.1183 -0.1320 0.0165 -0.0800  0.5728
[2,]  0.0987 1.0000 0.1928  0.0160  0.3530 0.1317  0.0955  0.2280
[3,]  0.1331 0.1928 1.0000  0.1871  0.0746 0.3632  0.0273  0.2147
[4,] -0.1183 0.0160 0.1871  1.0000  0.4128 0.4386  0.0952 -0.1638
[5,] -0.1320 0.3530 0.0746  0.4128  1.0000 0.2542  0.2274 -0.1492
[6,]  0.0165 0.1317 0.3632  0.4386  0.2542 1.0000  0.0707  0.0361
[7,] -0.0800 0.0955 0.0273  0.0952  0.2274 0.0707  1.0000  0.0417
[8,]  0.5728 0.2280 0.2147 -0.1638 -0.1492 0.0361  0.0417  1.0000

Median:           3.1352 104.3811 68.3522 19.2381 45.9682 29.5276 0.4185 30.0064

Mean:             3.298  109.98   68.184  19.664  68.792  30.3042 0.4297 31.19
MCD-estimated:
MDC-0.975-Mean:   2.1938 106.7656 68.4969 21.2531 66.4313 30.375  0.3921 25.3156
MDC-0.750-Mean:   2.1759 106.9198 68.6636 21.4444 67.9506 30.3611 0.395  25.4198
MDC-0.500-Mean:   2.1651 107.1371 68.6324 21.2804 67.2523 30.3648 0.3941 25.2804

Measures:
Mah.Dist:                       1.3823
Mah.Dist-MCD-0.975:             2.3485
Mah.Dist-MCD-0.750:             2.3532
Mah.Dist-MCD-0.500:             2.3541
```