1. Title: 1985 Auto Imports Database

2. Source Information:
   -- Creator/Donor: Jeffrey C. Schlimmer (Jeffrey.Schlimmer@a.gp.cs.cmu.edu)
   -- Date: 19 May 1987
   -- Sources:
      1) 1985 Model Import Car and Truck Specifications, 1985 Ward's
         Automotive Yearbook.
      2) Personal Auto Manuals, Insurance Services Office, 160 Water
         Street, New York, NY 10038
      3) Insurance Collision Report, Insurance Institute for Highway
         Safety, Watergate 600, Washington, DC 20037

3. Past Usage:
   -- Kibler,~D., Aha,~D.~W., \& Albert,~M. (1989). Instance-based prediction
      of real-valued attributes. {\it Computational Intelligence}, {\it 5},
      51--57.
         -- Predicted price of car using all numeric and Boolean attributes
         -- Method: an instance-based learning (IBL) algorithm derived from a
            localized k-nearest neighbor algorithm. Compared with a
            linear regression prediction...so all instances
            with missing attribute values were discarded. This resulted with
            a training set of 159 instances, which was also used as a test
            set (minus the actual instance during testing).
         -- Results: Percent Average Deviation Error of Prediction from Actual
            -- 11.84% for the IBL algorithm
            -- 14.12% for the resulting linear regression equation

4. Relevant Information:
   -- Description
      This data set consists of three types of entities: (a) the
      specification of an auto in terms of various characteristics, (b)
      its assigned insurance risk rating, (c) its normalized losses in use
      as compared to other cars. The second rating corresponds to the
      degree to which the auto is more risky than its price indicates.
      Cars are initially assigned a risk factor symbol associated with its
      price. Then, if it is more risky (or less), this symbol is
      adjusted by moving it up (or down) the scale. Actuarians call this
      process "symboling". A value of +3 indicates that the auto is
      risky, -3 that it is probably pretty safe.

      The third factor is the relative average loss payment per insured
      vehicle year. This value is normalized for all autos within a

particular size classification (two-door small, station wagons, sports/speciality, etc...), and represents the average loss per car per year.

-- Note: Several of the attributes in the database could be used as a "class" attribute.

5. Number of Instances: 205

6. Number of Attributes: 26 total
-- 15 continuous
-- 1 integer
-- 10 nominal

7. Attribute Information:

Attribute:	Attribute Range:
1. symboling:	-3, -2, -1, 0, 1, 2, 3.
2. normalized-losses:	continuous from 65 to 256.
3. make:	alfa-romero, audi, bmw, chevrolet, dodge, honda, isuzu, jaguar, mazda, mercedes-benz, mercury, mitsubishi, nissan, peugot, plymouth, porsche, renault, saab, subaru, toyota, volkswagen, volvo
4. fuel-type:	diesel, gas.
5. aspiration:	std, turbo.
6. num-of-doors:	four, two.
7. body-style:	hardtop, wagon, sedan, hatchback, convertible.
8. drive-wheels:	4wd, fwd, rwd.
9. engine-location:	front, rear.
10. wheel-base:	continuous from 86.6 120.9.
11. length:	continuous from 141.1 to 208.1.
12. width:	continuous from 60.3 to 72.3.
13. height:	continuous from 47.8 to 59.8.
14. curb-weight:	continuous from 1488 to 4066.
15. engine-type:	dohc, dohcv, l, ohc, ohcf, ohcv, rotor.
16. num-of-cylinders:	eight, five, four, six, three, twelve, two.
17. engine-size:	continuous from 61 to 326.
18. fuel-system:	1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi.
19. bore:	continuous from 2.54 to 3.94.
20. stroke:	continuous from 2.07 to 4.17.
21. compression-ratio:	continuous from 7 to 23.
22. horsepower:	continuous from 48 to 288.
23. peak-rpm:	continuous from 4150 to 6600.

24. city-mpg: continuous from 13 to 49.
25. highway-mpg: continuous from 16 to 54.
26. price: continuous from 5118 to 45400.

8. Missing Attribute Values: (denoted by "?")

Attribute #:	Number of instances missing a value:
2.	41
6.	2
19.	4
20.	4
22.	2
23.	2
26.	4