**"Дейтасет Boston Housing"**

Този набор от данни съдържа информация, събрана от Управлението на САЩ за населението, отнасяща се до жилищата в района на Бостън, Масачузетс. Тя е получена от архива на StatLib (http://lib.stat.cmu.edu/datasets/boston) и е широко използвана в литературата за оценка на алгоритми. Данните са първоначално публикувани от Харисън, Д. и Рубинфелд, Д.Л. "Хедонични цени и търсенето на чист въздух", J. Environ. Economics & Management, том 5, стр. 81-102, 1978 г.

Наборът от данни за жилищата в Бостън е малък, особено в днешната ера на големите данни. Всяка от 506-те реда в набора от данни описва предградие или град в Бостън и има 14 колони с информация, като среден брой стаи на жилище, съотношение ученици-учители и престъпност на глава на населението. Последният ред описва медианната цена на домовете, заети от собственици (без да включва домове, които се отдават под наем), и обикновено този ред е този, който се опитваме да предвидим, когато го използваме за задачи по регресия. Но заедно с това може да се предскаже и нивото на азотен оксид

crim: Това е нивото на престъпност на града на глава на населението.

zn: Това е дялът на земята за жилища, зонирана за парцели с площ над 25,000 кв. фута.

indus: Това е дялът на неразнищни търговски акри на града.

chas: Това е фиктивна променлива за река Чарлс (= 1, ако районът граничи с реката; 0 в противен случай).

nox: Това е концентрацията на азотни оксиди (части на 10 милиона).

rm: Това е средният брой стаи на жилище.

age: Това е дялът на обитаваните единици, построени преди 1940 г.

dis: Това са утеглени разстояния до пет бостънски центъра за заетост.

rad: Това е индекс на достъпност до радиални магистрали.

tax: Това е ставката на данък върху имотите с пълна стойност за всеки $10,000.

ptratio: Това е съотношението между учениците и учителите по градове.

b: Това се изчислява като 1000(Bk – 0.63)^2, където Bk е дялът на хора с афроамерикански произход по градове.

lstat: Това е процентът на хората с по-нисък социален статус.

medv: Това е медианната цена на къщите, заети от собственици, в хиляди долара и често е целевата променлива в задачи по регресия.