Supplementary material: Bodyfat estimation data set

The *Bodyfat Estimation* dataset can serve as an example for a real world regression problem, see for instance http://garthtarr.github.io/mplot/reference/bodyfat.html
The dataset comprises 13-dim. feature vectors which represent the following quantities or measurements in 252 subjects:
age in years
weight in kg
height in inches
neck circumference in cm
chest circumference in cm
abdomen circumference in cm
hip circumference in cm
thigh circumference in cm
knee circumference in cm
ankle circumference in cm
biceps circumference in cm
forearm circumference in cm
wrist circumference in cm

The target of the regression is to predict the bodyfat percentage. Target values **T** range from 0 to 50, while the range of observed values is different for each of the 13 features. Original data is provided in the folder as “originaldata.mat” for easy import into matlab. Alternatively, files “torig.txt” (targets) and “xorig.csv” (features) are provided as well as *.txt versions of the files.

For the analysis along the lines of Assignment III (soft-committee machine, bonus suggestions) pre-processed versions of the data have been prepared:

“xzscore.csv” : z-score transformed feature vectors, where each transformed feature displays zero mean and unit variance over the 252 samples
“tshift.csv” : shifted/rescaled target values t = (T- mean(T))/30 resulting in a range of -1 < t < 1 and mean(t)=0.

Alternatively the file “transformeddata.mat” contains both the transformed feature vectors and target values for easy matlab import.

As a potential bonus problem w.r.t. assignment 3, one could split the (transformed) data randomly in, for instance, 200 training samples and 52 validation or test samples. Training of a soft-committee with, say, K=5 and adaptive weights v_k should yield non-trivial performance in terms of the MSE cost function.