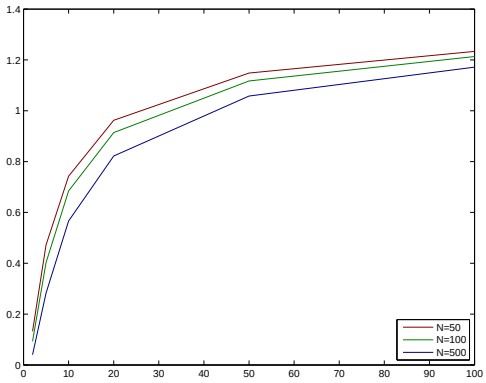
Problem Set 1
Solutions

Question 1.

```
% script q1 to answer question 1
ds = [2 5 10 20 50 100];
m = 100;
h= plotavemin(50,m,ds);
set(h,'Color',[0.5 0 0]);
hold on;
h= plotavemin(100,m,ds);
set(h,'Color',[0 0.5 0]);
h= plotavemin(500,m,ds);
set(h,'Color',[0 0 0.5]);
legend('N=50','N=100','N=500','Location','SouthEast');
hold off;
```




```
function h = plotavemin(n,m,ds)
% function h = plotavemin(n,m,ds)
% plots the average minimum distance as a function of the
% number of dimensions for points randomly drawn from the
% unit ball.
% n is the number of points, m is the number of trials,
% ds is a vector of the number dimensions to try
% h is the handle to the plotted line

dists = zeros(size(ds));
i = 1;
for d = ds
    for j=1:m
        % draw n d-dim pts at random from unit hyperball:
        pts = randn(n,d);
        pts = pts.*repmat((rand(n,1).^(1/d))./sqrt(sum(pts.*pts,2)),[1 d]);
        % calculate squared distances between all pairs:
        D = repmat(pts,[1 1 n]) - repmat(permute(pts,[3 2 1]),[n 1 1]);
        % second term to make diag elements large (so pts isn't its own NN):
        D = sum(D.*D,2) + permute(eye(n,n),[1 3 2])*10;
        % find average minimum distance and add it to running sum:
        dists(i) = dists(i) + mean(sqrt(min(D,[],3)));
    end;
    dists(i) = dists(i)/m;
    i = i+1;
end
h = plot(ds,dists);
```

The maximal distance between any two points is 2. The fewer the points, or the higher the dimensionality of the space, the farther apart a point is from its closest neighbor. That means that all of the points are approximately equally far away if the dimensionality is high enough (all are near a distance of 2). For function approximation to work, we need to exploit the locality of the space (or the smoothness of the function). If there is no notion of a "neighborhood" then function approximation breaks down. Figure 2.6 shows something similar: most of the volume is far away in high dimensional spaces. Equation 2.24 demonstrates the same as shown here, but for the median (instead of the mean).