**Second Assignment** (see also slide 91 of the ‘slides of lecture 1-3’, available on BlackBoard)

* Assignment description (in groups of 1 to 3 persons):
  1. Assume a data generating process data points (**x,** y), where samples of **x** = (*x*1, *x*2, …, *xp*) are generated *uniformly* on unit cube [0,1]*p* and related *y*-values are calculated according to *y* = exp (-a||**x||**), so without any measurement error; a is a parameter, the precise value of which you may choose yourself (not too big and not too small…), *p* equals the dimension of unit cube [0,1]*p*.  
     Note that ||**x**|| = sqrt (x12 + x22 + ... + x*p*2) , i.e. the Euclidian length of vector **x**  (see also [http://en.wikipedia.org/wiki/Euclidean\_distance](https://webmail.tudelft.nl/owa/redir.aspx?C=AOtk41Pk3UyCVlueFyJxBIbYJpdlY9IIntGu47grkitYEjBziFGG91cMgIeT8Qvs7qv54QR_lgM.&URL=http%3a%2f%2fen.wikipedia.org%2fwiki%2fEuclidean_distance)).
  2. Create a series of six data sets, *each* one of 1000 samples, according to the given specification, for values of *p* =1, 2, 4, 8, 16, 32, respectively.
  3. Use the 1-nearest-neighbor rule to *predict* the value of *y*0 at the test-point x0 = 0 (note that *y*0 = 1) and do this for all six data sets; given *p*, the *prediction* of *y*0 is denoted as *y*0 *p*.  
     Note that for all predictions: *y*0 *p* <= 1.
  4. Create a figure showing the six predictions *y*0 *p* as function of *p:* what do you conclude? Can you link the results found with the formula *ep* (*r* ) = *r* 1*/p* as given on slide 83of ‘slides of lecture 1-2’?
  5. Repeat the experiment a few times to analyze sensitivity to the data generation process.
  6. Define the *prediction error* as *errp* = 1 - *y*0 *p*. In order to keep *errp  approximately* *the same* while enhancing *p*, a growing number of sample points is needed. Assuming again uniformly distributed sample points on unit cube [0,1]*p* and starting with a self-chosen (relatively small) sample size *sp* for

*p* = 1, what would be your guess about the required sample size *sp* as function of *p* in order to keep *errp  approximately* *the same*? Design and execute a simulation (= an experiment) that tests your guess.

* 1. Write a **short report** describing the findings of your work w.r.t. this assignment 2: describe motivation, set up of experiments, results and conclusions; submit your report **both in hard and soft copy** (hard copy can be delivered at the start of lecture or in my physical postbox at TPM (in wing B, floor 3 of building 31).
  2. **Deadline**: **May 18, 10:45am, 2016.**