Answers to questions about part 1

1) The reason tiles in the second tiling run from 121 to 241 is to disambiguate what tiling we are on (1 - 8) and where in that tiling we are (0 - 120 + offset).
2) It will be in the first tile in the first 7 tilings since as we slide the tile, for the first 7 tiles only, (0.1, 0.1) is going to be contained in the bounding box that is formed by the bottom left and top right corners.
3) The 8th tilling does not contain the point in its first tile because it exceeds the bounding box defined by it. The point (0.1, 0.1) lies in the bounding box of the 13 tile.
4) Since we know it lies within the 13th tile (12th index) of the 8th tiling, we know it is 12 + (7 * 121) = 859.
5) We know that the last possible index you can reach is 121 (120th index), therefore using our formula 120 + (7 * 121) will get us our last possible index which is equal to 967.
6) The difference between in1 is 0, and in2 is only off by .1. Relative to how much the tiles are translated, a difference of 0.1 is not a large difference, therefore most tile indices should be the same.

Answers to questions about part 2

1) Example 4’s before is non-zero because it shares the majority of tilings with example 2.
2) It does not decrease further to zero since MSE is measuring the difference between the actual value and our approximate value using tiling. In order to make a continuous space discrete we have to find a compromise between accuracy and efficiency.

Explanation of part 2

There is one tall bimodal peak, one deep valley and a bunch of smaller valleys. Starting with the peak, its tallest point resides at about (1.8, 3.0) with a height of about 0.1. The area surrounding the the bimodal peak (1.8, 3.0) is roughly 2.5 by 1.5. The target function at this point (sin(in1 - 3) * cos(in2) + N(0, 0.1)) is positive so our estimate is modeling that. Now for the large valley, it is located at (4.3, 2.9) with a height of -0.1. Its area spans 1 by 1, and the target function at this point is reflected. Finally the smaller peaks and valleys are the intermediate values between the max and min of the target function.

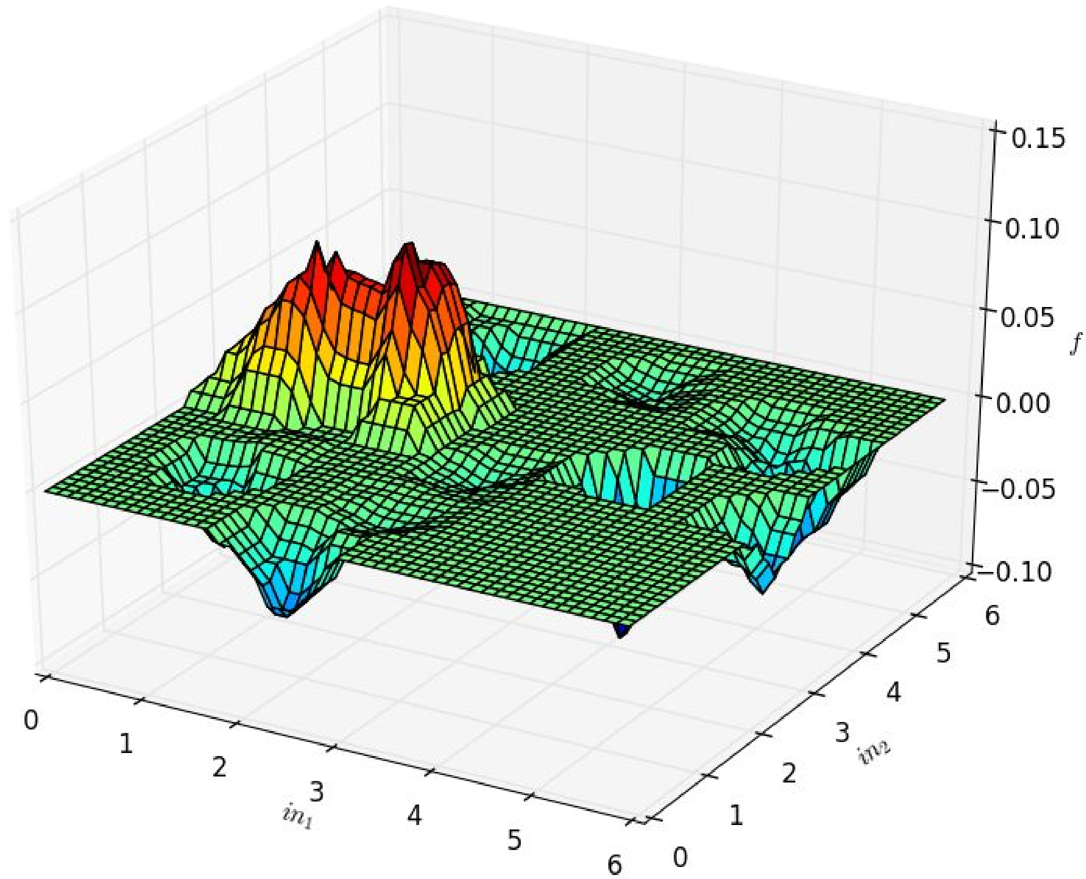
As a continuous space is approached, the height at (in1, in2) will start to look more like small pins, especially with a small number of inputs to learn from. As our sample size grows, a tiling window with a lot of tiles will have a lot more accuracy than the tiling with only a few tiles. Considering this, if we divide in1 to become more fine grained then we will get a jagged representation in its axis while a more smoothed out representation in the other, for a small sample size.

MSEs

Example (0.1 , 0.1 , 3.0): f before learning: 0.0 f after learning : 0.3
Example (4.0 , 2.0 , -1.0): f before learning: 0.0 f after learning : -0.1
Example (5.99 , 5.99 , 2.0): f before learning: 0.0 f after learning : 0.2
Example (4.0 , 2.1 , -1.0): f before learning: -0.075 f after learning : -0.1675

The estimated MSE: 0.251952360461
The estimated MSE: 0.0513023349587
The estimated MSE: 0.0200579042604
The estimated MSE: 0.0143183207319
The estimated MSE: 0.0125313139164
The estimated MSE: 0.01153070221
The estimated MSE: 0.0115925131493
The estimated MSE: 0.0111132080627
The estimated MSE: 0.011192041522
The estimated MSE: 0.0114058165732
The estimated MSE: 0.0111609122467

F20



F10000