Answers to questions about part 1

1. Since every tiling needs specific tile number and the first tiling used 0 -120 (11*11 = 121 tiles), the second tiling starts its 121 tiles from 121 to 241.

2. (0.1, 0.1) has ((-0.6/8) * tiling_num, (-0.6/8) * tiling_num) offset depending on the tile number. The first one, for instance, has (0, 0) offset and the seventh one has (0.45, 0.45). Since after this offset is added to the original point (0.1, 0.1), it would not be more than 0.6 in both in1 and in2 directions. Therefore, (0.1, 0.1) will be on the first tile of the first seven tilings.

3. The eighth tiling padding will be (0.525, 0.525) and as it is added to the point (0.1, 0.1), it will exceed 0.6 in both in1 and in2 directions. Hence, it will be moved one tile in both directions and that will be on thirteenth tile of the eights tiling.

4. The first tile of the eighth tiling is 121*(8-1) + (1-1), thus the thirteenth tile will be 121*(8-1) + (13 - 1), which is 859.

5. As each tiling has 121 (11*11) number of tiles, the last tile of the eighth tiling will be on 121*(8-1) + (121 - 1), which is 967.

6. Points (4, 2) and (4, 2.1) are very close to each other. Since the offset for each tiling is ((-0.6/8) * tiling_num, (-0.6/8) * tiling_num), that (0, 0.1) difference between those two points would not make much of a difference in the set of indices. In addition, this is exactly what we want for the purpose of generalization.

Explanations of part 2

The learned function with 20 samples is not close to the target function because of the few number of samples that can not cover the input space. Since the initial value for the approximated function is 0, most of it is still 0 as there were no samples to cover those areas.

There are approximately 12 peaks and valleys, which five of them are sharp due to exaggeration made by the few number of samples in those areas. Smoother peaks and valleys happened as we got more samples close to them.

Since peaks and valleys could not be smoothed due to few number of samples, the width of them are approximately 1 tile.

If we use 11x21 tiling, since we had 21 tiles for in1 direction, peaks and valleys would be thinner in in1 direction. However, it would be the same for in2 direction as we did not change tiling number for in2 direction. In addition, as the number of samples is still very low to approximate the target function, we would see peaks and valleys with the width of 1 tile.

Answers to questions about part 2

The forth sample, (4.0, 2.1), has 6 out of 8 tiles in common with the second sample (4.0, 2.0), which causes the before value for the forth sample to be non-zero. In fact, it is 6/8 * after_value of the second sample, which is 6/8 * (- 0.1) = -0.075.

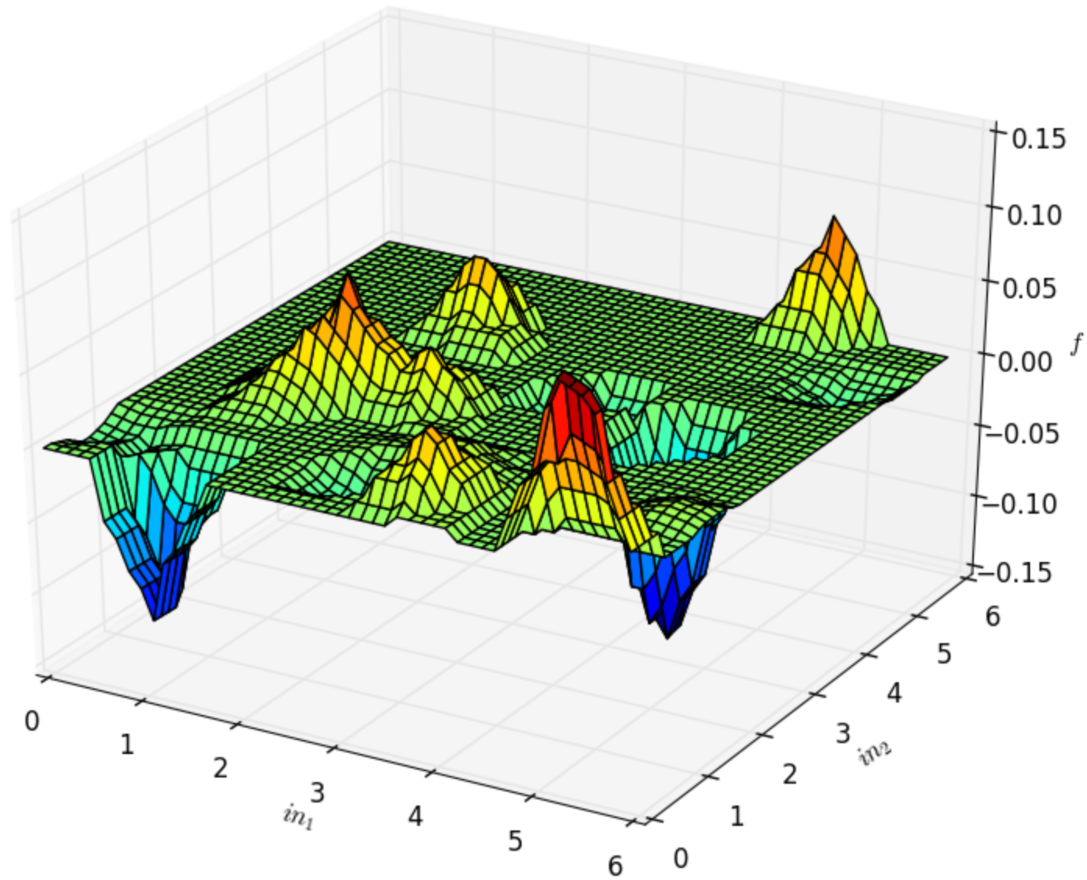
Since the target includes a random number with normal distribution and standard deviation of 0.1, the estimated MSE will not decrease towards zero after hitting around 0.01. In fact, as the standard deviation is 0.1 and we use mean squared error, since the square of 0.1 is 0.01, we can not get any better approximation.

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: 0.24958134153
The estimated MSE: 0.0508037360229
The estimated MSE: 0.0207391367403
The estimated MSE: 0.0147958662403
The estimated MSE: 0.0126347538314
The estimated MSE: 0.0120276274072
The estimated MSE: 0.0117809247323
The estimated MSE: 0.0113391643061
The estimated MSE: 0.0112162930193
The estimated MSE: 0.0110867423553
The estimated MSE: 0.0111461885051

F20



F10000