Marcin Pietrasik
Sean Scheideman
Explain

The learned function after 20 examples looks like a plane with several peaks and valleys. The reason it looks like this is because the function has only began learning. From the 20 points it was given, it doesn't have enough information about all of the points to create an accurate model of the target function. No points, for example, fell in range x=[0,2] and y=[4,6], this is why the learned function in the range is unchanged (f(x,y) = 0). There are about 16 distinct peaks and valleys. This number is lower than the 20 points because it is possible that there were two or more identical or very similar points, where the peaks are blended together. The height of the peaks and valleys depends the height of the target function, if the target function is very low or very high, the learned weights will reflect that. Also, if there have been points with shared tilings, this will create more pronounced peaks and valleys. Finally, the error between the target and the inner product is multiplied by the step size parameter (0.0125) so even though the peaks and valleys are moving closer to the target, there have not been enough updates to move it near the target function. The widths of the peaks and valleys are affected by the proximity of the 20 points. If there are many points close together, they will create a wide peak or valley. In addition the width will be affected by the size of the tiles used (0.6X0.6) and the amount of offset of each tiling (0.075), so a single point that is passed to learn() will impact the approximated function in an oval shaped area with a positive 0.5 slope major axis.

If asymmetric generalization was used where the x axis was divided twice as finely as before, the learned function after 20 examples would have peaks and valleys that cover an oval shaped area with a narrower minor axis as opposed to square tilings. Also, the peaks and valleys along the x axis would be sharper due to the greater precision in the tilings along the x axis.