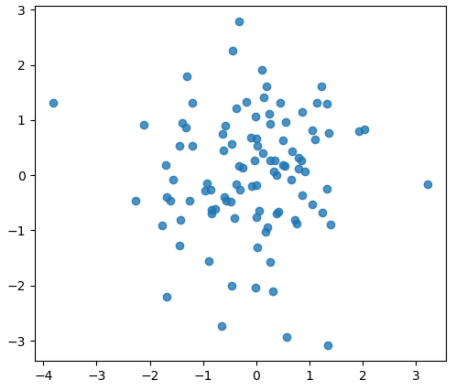
Progress Report
Model Setup
March 6, 2018

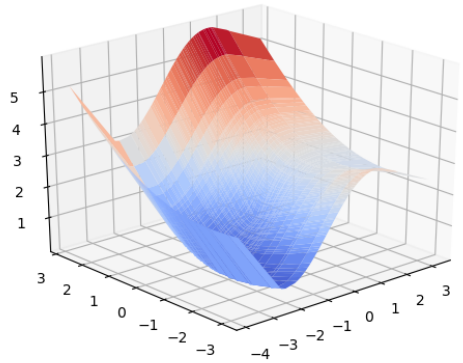
Simulated Data

First generate data from a spatial gaussian process with 100 locations, we sample spatial locations (x, y) iid from standard normal, and assume the pollutant z follow below Gaussian Process:

$$z(x,y) \overset{iid}{\sim} N(f(x,y), \sigma^2 = 0.1)$$
$$f(x,y) = 0.2x + 0.5y + \sqrt{x^2 + y^2} + sin(x) + cos(x)$$



(a) Sampled saptial location for monitoring sites (standardized)



(b) Average pollutant surface over space (standardized)

We then generate prediction for $z(x, y)$ from 5 base GP models, with covariance structure:

1. Linear,
2. Polynomial, degree 3
3. Gaussian RBF, with ARD
4. Matérn $\frac{5}{2}$, with ARD
5. MLP, with ARD.
   Equivalent to a 2-layer network with Gaussian CDF activation function and infinite hidden units:

$$k(x,y) = \sigma^2 \frac{2}{\pi} \text{asin}\left(\frac{\sigma_w^2 x^\top y + \sigma_b^2}{\sqrt{\sigma_w^2 x^\top x + \sigma_b^2 + 1}\sqrt{\sigma_w^2 y^\top y + \sigma_b^2 + 1}}\right)$$

The out-of-sample MSE for 5 models are 4.846, 2.094, 2.011, 1.989, 1.981 respectively.