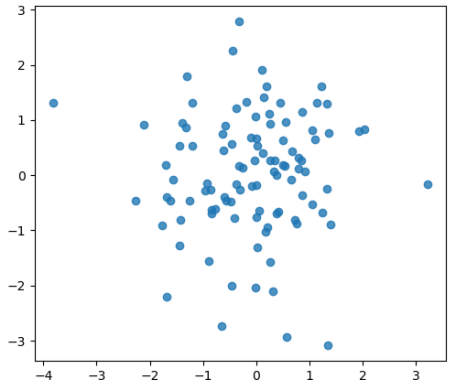
Progress Report
Model Setup
March 6, 2018

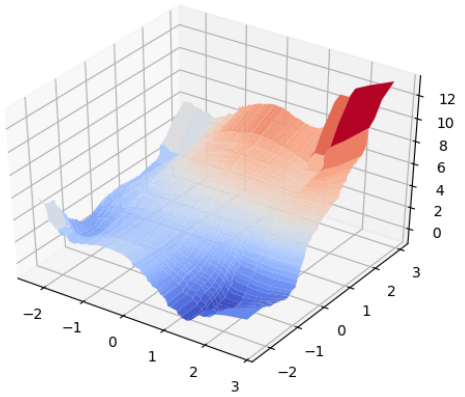
## Simulated Data

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.

## Model

Given data $\{(x_i, y_i), z_i\}_{i=1}^{n}$ and $b \in \{1, \ldots, B\}$ base models, the spatial-temporal ensemble model can be expressed in terms of below probablistic program.

$$
\begin{aligned}
z &= \sum_{b=1}^{B} w_b(x,y) * z_b(x,y) \\
\mathbf{w} &= softmax(\mathbf{w}') \qquad \text{where } \mathbf{w} = \{w_b\}_{b=1}^{B}, \mathbf{w}' = \{w'_b\}_{b=1}^{B} \\
w'_b(x,y) &\sim \mathcal{GP}(0, \quad k_{\boldsymbol{\theta}}(x,y) + \sigma^2), \quad \boldsymbol{\theta} \sim P() \\
z_b(x,y) &\sim \mathcal{GP}(\hat{z}_b(x,y), \quad cov(z_b)(x,y))
\end{aligned}
$$

Notice that the distribution of $z'_b s$ are fixed *a priori* (already learned from pre-processing step), and inference is performed for $\mathbf{w}_b$ along.
This model will be implemented in pymc3 language.

## Timeline

1. (March. Week 3-4)
   Initial implementation.
2. (March. Week 4 - April. Week 1)
   Initial experiment on simulated data. Investigating:
   (a) Sensitivity to model and kernel specification
   (b) Whether improved performance for overall ensemble, how it is tighed to the property of the base models. More specifically:
      i. Variance among model prediction
      ii. Number and bias of models (few strong model with good prediction for different aspect of the data, or large collection of weak models)
   (c) Identifiability of individual weights
3. (April. Week 2-3)
   More realistic experiment, gradually increase complexity of data-generation mechanism toward (using the mean-surface from QD or Itai's prediction)
4. (April. Week 4) Start experiment on real data