Probabilistic Numeric Convolutional Neural Networks

(Finzi et al 2020)

Discussion Leader:

Shah Rukh Qasim

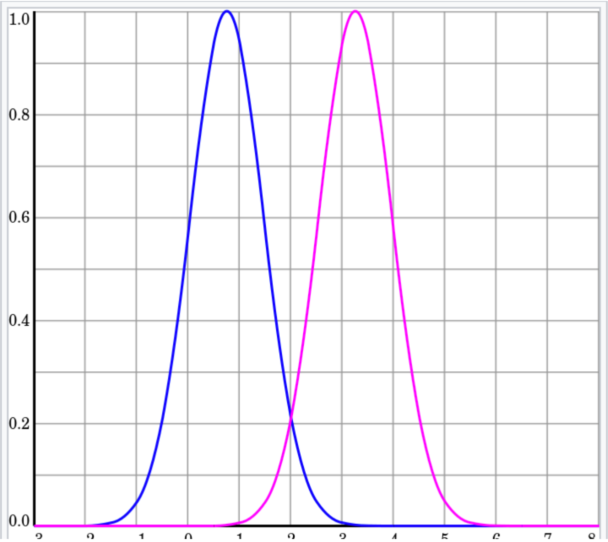
Introduction

- Essentially: Extension of **CNN**s to **continuous signals** that are **irregularly sampled** (I can’t help but think about HGCal) done using **Gaussian Processes (GP)** to fill the gaps and numerical methods for approximations (and proofs?).
- Dataset used (one of them): Superpixel MNIST is an adaptation of the MNIST dataset where the 784 pixels of the original images are replaced by 75 salient *superpixels* that are non uniformly spread throughout the domain and are different for each image (Monti et al., 2017).
- Imagine HGCal data or any calorimetric data: Is the underlying signal continuous?

Gaussian Processes

- Let’s start with a nice animation here:
- https://distill.pub/2019/visual-exploration-gaussian-processes/

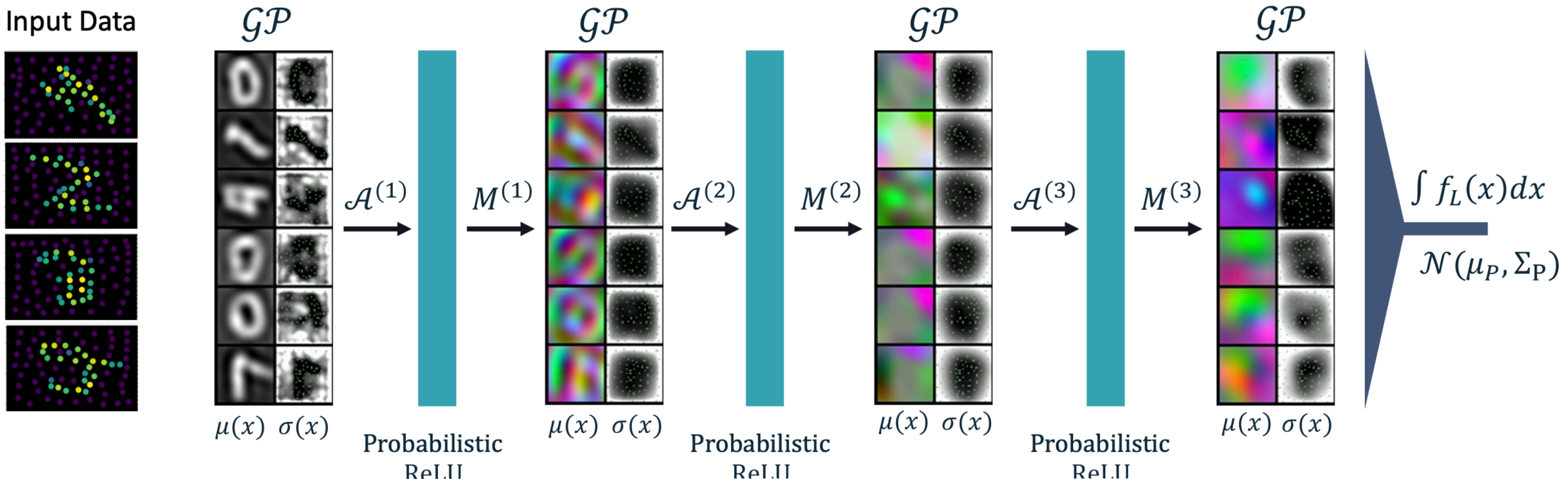
- Done using covariance, mean matrix
- Simple calculation using cov/mean matrix



Source: wikipedia

$$\mu_p(x) = \boldsymbol{k}(x)^\top [K + S]^{-1} \boldsymbol{y}\,, \quad k_p(x, x') = k(x, x') - \boldsymbol{k}(x)^\top [K + S]^{-1} \boldsymbol{k}(x')$$

Hollistic view



$$f^{(\ell+1)} = M^{(\ell)}\mathrm{ReLU}[\mathcal{A}^{(\ell)} f^{(\ell)}],$$

$$\mathcal{P}(f^{(L)})_\alpha = \int f_\alpha^{(L)}(x)\mathrm{d}x \text{ for each } \alpha = 1, 2, \ldots, c.$$

- Discuss HGCal data again
- Challenge with windowing
- Can we take the whole calorimeter?

PNCNNs

$$f^{(\ell+1)} = M^{(\ell)}\mathrm{ReLU}[\mathcal{A}^{(\ell)} f^{(\ell)}],$$

$$\mathcal{A} = \sum_k W_k \mathrm{e}^{\mathcal{D}_k}$$

- e^{D_k} is diffusion equation
- For discrete case it can be defined as

$$\mathcal{D}_k = \beta_k^\top \nabla + \frac{1}{2}\nabla^\top \Sigma_k \nabla$$

- [Page 4 of the paper]

Application to radial basis kernal (RBF)

- Closed form solution of mean/variance
- Then reduces to simple convolution

- Discuss from the paper [Page 5]

Channel mixing

- After non-linearity, can’t go with gaussian process assumption
- Thanks to central limit theorem, after mixing channels, you get gaussian assumption back

$$\mathbb{E}[f^{(\ell+1)}(x)] = M\mathbb{E}[h^{(\ell)}(x)], \quad \mathbb{E}[f^{(\ell+1)}(x)f^{(\ell+1)}(x')^\top] = M\mathbb{E}[h^{(\ell)}(x)h^{(\ell)}(x')^\top]M^\top$$

Training procedure

- Channel mixing, diffision: $\{(M^{(\ell)}, \boldsymbol{W}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \boldsymbol{\Sigma}^{(\ell)})\}_{\ell=1}^{L},$
- Hyperparameters of kernel $\{(l^{(\ell)}, a^{(\ell)})\}_{\ell=1}^{L}$
- Loss is $L_{\text{task}} + \lambda L_{\mathcal{GP}}$
- [Go to paper page 7, equation 11 to discuss exact loss function]

Final Remarks

- Should definitely be looked into
- We can start with HGCal data with single particle shot into it and try to regress a class or energy
- Or the toyset
- Code available