Logo

Uncertainty propagation in neural networks for sparse coding

Danil Kuzin, Olga Isupova, Lyudmila Mihaylova

University of Sheffield, UK, University of Oxford, UK

1-LISTA

Reconstruct $\boldsymbol{\beta}$ from observations $\mathbf{y}$ collected as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, such that elements $\boldsymbol{\beta}$ contain zeros.

- Represent iterative soft-thresholding algorithm as a recurrent neural network with shared weights
- Learn weights with backpropagation through time
- Overfitting
- No uncertainty estimation

Require: observation $\mathbf{y}$, current weights $\mathbf{W}, \mathbf{S}$, number of layers L

1: *Initialisation.* Dense layer $\mathbf{b} \leftarrow \mathbf{W}\mathbf{y}$
2: *Initialisation.* Soft-thresholding nonlinearity $\widehat{\boldsymbol{\beta}}_0 \leftarrow h_\lambda(\mathbf{b})$
3: **for** $l = 1$ **to** L **do**
4: Dense layer $\mathbf{c}_l \leftarrow \mathbf{b} + \mathbf{S}\widehat{\boldsymbol{\beta}}_{l-1}$
5: Soft-thresholding nonlinearity $\widehat{\boldsymbol{\beta}}_l \leftarrow h_\lambda(\mathbf{c}_l)$
6: **end for**
7: **return** $\widehat{\boldsymbol{\beta}} \leftarrow \widehat{\boldsymbol{\beta}}_L$

3-Uncertainty propagation

At every step the output of soft-thresholding can be closely approximated with spike and slab distribution

1. $\mathbf{b} = \mathbf{W}\mathbf{y}$ is Gaussian-distributed
2. $\widehat{\boldsymbol{\beta}}_0 = h_\lambda(\mathbf{b})$ is approximated with the spike and slab distribution
3. $\mathbf{e}_l = \mathbf{S}\widehat{\boldsymbol{\beta}}_{l-1}$ is approximated with the Gaussian distribution
4. $\mathbf{c}_l = \mathbf{b} + \mathbf{e}_l$ is Gaussian-distributed
5. $\widehat{\boldsymbol{\beta}}_l = h_\lambda(\mathbf{c}_l)$ is approximated with the spike and slab distribution

- All latent variables are modelled with parametrised distributions
- We can apply approximate Bayesian inference methods

2-BayesLISTA

- Add priors for NN weights
- Perform full Bayesian inference
- Estimate uncertainty of predictions

4-BackProp-PBP

Blocktext

5-Results

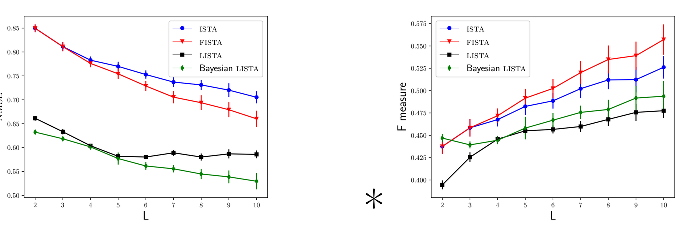

Fig. 1: Different depth performance

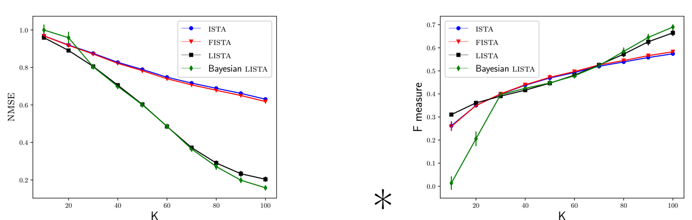

Fig. 2: Different observation size performance

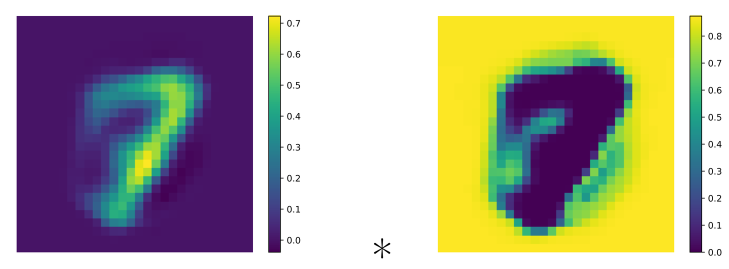

Fig. 3: Posterior parameters for an image of digit 7

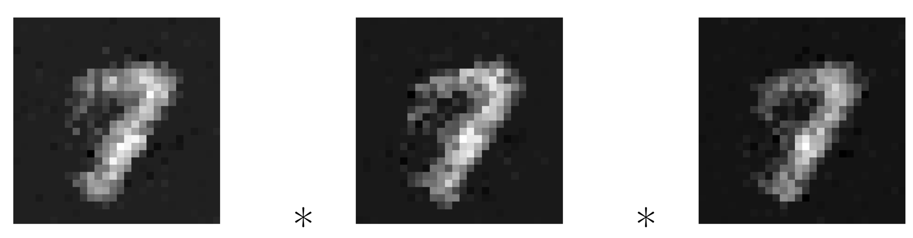

Fig. 4: Samples from the posterior for an image of digit 7

Use the estimated uncertainty to choose next training data with largest variance

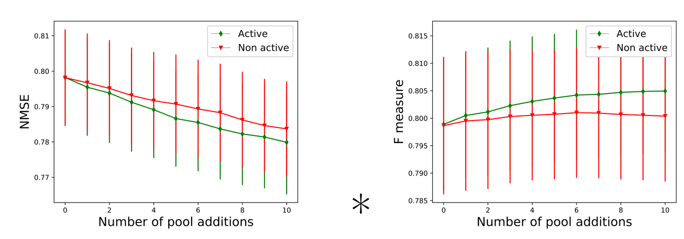

Fig. 5: Sequential pool additions

6-Summary

Blocktext

Notetext