Mathematical Foundation of Machine Learning – Spring 2018

Project list

Theoretical projects

1. [Anisotropic Besov smoothness] Let $f(x) = I_{\tilde{\Omega}}(x)$, where $\tilde{\Omega} \subset [0,1]^3$ is convex with smooth boundary. Let $p = 2$. Show that
   a. For a tree $\mathcal{T}_D$, created via non-adaptive partition along dyadic lines, we have $f \in B_\tau^\alpha(\mathcal{T}_D)$, for $\alpha < 1/4$ and $1/\tau = \alpha + 1/2$.
   b. Show there exists an adaptive anisotropic tree $\mathcal{T}_A$, such that $f \in B_\tau^\alpha(\mathcal{T}_A)$, for $\alpha < 1/2$ and $1/\tau = \alpha + 1/2$.

2. [Scattering Networks] Scattering Networks are a special (simplified) form of convolutional neural networks where the coefficients of the filters predetermined and not learnt. Provide an overview of the method [6] and in particular, a sketch of the "stability to deformation" property [5].

Applied projects

General comments

A. Use publically available datasets such as the UCI machine learning repository (http://archive.ics.uci.edu/ml/index.php)
B. In all experiments use 5 fold cross validation.
C. For regression problems provide average error and std of error statistics.
D. For classification problems provide accuracy (TP+TN)/(P+N), precision TP/(TP+FP) and recall TP/(TP+FN) statistics.
E. Perform hyper-parameter search and try to explain the logic of the best configuration.
F. "Debug" your results: look at confusion matrices, investigate your false positive and negatives. Try to understand where your models fail and try to fix them.
G. For ML problems, compare your results to the results using standard models from Scikit-Learn, R, etc.
H. Try to come up with other ideas beyond the basic project description.

3. [Feature importance via wavelet decomposition of RF] Reproduce the feature importance results of [1]
   a. Provide summary of the wavelet-based method with emphasis on the use of the validation set to determine a threshold for wavelet norms.
   b. Test on regression & classification problems (multi-class problems).
   c. Compare extensively with standard methods as in [1]

4. [Compression & denoising with wavelet decomposition of RF] Reproduce and add to the results of [1]
    a. Compression – Investigate the RF compression capabilities of wavelets through tradeoff between the number of trees and tree components versus the prediction error.
    b. Denoising – add various levels of Gaussian noise to regression datasets and add various levels of mis-labeling to classification datasets. Investigate the performance of wavelet denoising.

5. [Smoothness analysis of DL] Reproduce the research of the paper [3]
    a. Train the 'classical' TensorFlow networks for MNIST, CIFAR10 datasets.
    b. Perform Besov smoothness analysis of the representation layers at various stages of the training
    c. Try also to experiment with different network configurations and investigate the relationship between the classification error on the testing set and the smoothness analysis of the representation layers.

6. [DL classification as regression] Add to the research of [3]
    a. Take the 'classical' TensorFlow networks for MNIST, CIFAR10 datasets and modify the last layer from "softmax" to linear mapping into $\mathbb{R}^9$.
    b. Assign each class a vertex on the standard simplex in $\mathbb{R}^9$.
    c. Define a loss function that minimizes the sum of squared distances to the corresponding vertex of the class of each sample.
    d. Test various configurations with this last layer (e.g. adding more convolution or fully connected layers before it) and benchmark with 'classical' architecture.
    e. Perform Besov smoothness analysis and correlate with your performance results.

7. [Analysis of ResNets] Add to the research of [3]
    a. Train a 'small' version of a ResNet [4] on the CIFAR10 dataset.
    b. Investigate the performance and apply smoothness analysis of the network with and without the residual connections.

8. [Anisotropic RF partitions] Solve regression problems using wavelet decompositions of anisotropic Random Forests
    a. Interface with the given platform to apply anisotropic hyperplane splits via linear regression and SVM.
    b. Compare with isotropic splits. In which cases did the anisotropic constructions give a better result? Why?
    c. Compute Besov smoothness of all the isotropic and anisotropic models and investigate if correlate with test results of the models.

9. [Multi-label problem via wavelet decomposition of GB] Extend the experimental results of [2] to the case of multi-label classification (more than 2 classes).

10. [Scattering Networks for small datasets] Use the methods of [6] to compare the results of trained convolutional neural networks over Scattering Networks in the cases of small/simple datasets.

a. Use the MNIST dataset as an example for a “simple” dataset.
b. Use various sizes of the CIFAR10 dataset.

References

[1] O. Elisha and S. Dekel, Wavelet decompositions of Random Forests - smoothness analysis, sparse approximation and applications, JMLR 17 (2016).

[2] O. Morgan, O. Elisha and S. Dekel, Wavelet decomposition of Gradient Boosting, preprint.

[3] O. Elisha and S. Dekel, Function space analysis of deep learning representation layers, preprint.

[4] H. Kaiming, Z. Xiangyu, R. Shaoqing and SD Jian, Residual Learning for Image Recognition, proceedings of CVPR 2016.

[5] S. Mallat, Group Invariant Scattering, Comm. Pure and Applied Math 65 (2012), 1331-1398.

[6] J. Bruna and S. Mallat, Invariant Scattering Convolution Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (2013), 1872 – 1886.