ACMS 80770-03: Deep Learning with Graphs
Homework 4

Handed out: Monday, November 28, 2022

Due: Thursday, December 8, 2022 10:00 PM

Notes:

- Software resources for this homework are available on course's GitHub repository. Only use the packages imported in the starter code for your implementations.
- **Don't fork the repository.** Download the code and add it to your current repository.
- Use the provided latex file to submit your written work.
- Submit your codes to Gradescope only through Bitbucket or GitHub.
- Upload your written work (PDF only) directly to Gradescope.
- Gradescope uses MOSS to detect similar submissions. Once detected, copying and pasting from classmates are regarded as a violation of the honor code.

Problem 1: Spectral wavelet transform consists of convolution of the signal $\boldsymbol{f} \in \mathbb{R}^{|V|\times d}$ residing on the graph $\mathcal{G} = (V, E)$ with a set of multi-scale spectral filters $\{\hat{g}_j\}_{j=1}^{\mathcal{J}}$ on different scales j,

$$\boldsymbol{f} *_{\mathcal{G}} g_j = \boldsymbol{\chi}\hat{g}_j(\boldsymbol{\Lambda})\boldsymbol{\chi}^T\boldsymbol{f}, \tag{1}$$

where $\boldsymbol{\chi}$ and $\boldsymbol{\Lambda}$ represent the eigenvector matrix and the diagonal matrix of eigenvalues of the normalized Laplacian matrix $\mathcal{L}_{sym}$, respectively.

Let $\hat{g}$ be a spectral window, translation and dilation of which constructs $\{\hat{g}_j\}_{j=1}^{\mathcal{J}}$ as

$$\hat{g}_j(\lambda) := \begin{cases} \hat{g}\left(\log(\lambda) - \frac{a(j-1)}{R}\right) & \text{for } 2 \le j \le \mathcal{J}, \\ \sqrt{Rd_0^2 + \frac{R}{2}\sum_{k=1}^{\mathcal{K}} d_k^2 - \sum_{i=2}^{\mathcal{J}} |\hat{g}_i(\lambda)|^2} & \text{for } j = 1, \end{cases} \tag{2}$$

where $a = \frac{R\log(\lambda_{\max})}{\mathcal{J}-R+1}$ is the dilation factor, R is the kernel overlap which tunes the filter's width in the spectral domain, and $\mathcal{J}$ is the number of the filters.

A. Write a function to compute the spectral filters $\{\hat{g}_j\}_{j=1}^{\mathcal{J}}$ constructed by the half-cosine kernel window

$$\hat{g}(\lambda) := \sum_{k=0}^{\mathcal{K}} d_k \left[\cos\left(2\pi k\left(\frac{\lambda}{a}+\frac{1}{2}\right)\right) \cdot \mathbb{1}\{0 \le -\lambda < a\}\right], \tag{3}$$

where $\mathcal{K} = 1$, $d_0 = d_1 = \frac{1}{2}$, $R = 3$, $J = 8$, and $\lambda_{\max} = 2$.

B. Plot the filters $\{\hat{g}_j\}_{j=1}^{\mathcal{J}}$ as a function of the λ in the interval $[0, \lambda_{\max}]$.

Problem 2: Graph Scattering Network is a graph convolutional neural network architecture that uses wavelet transform building blocks to generate invariant representation $\boldsymbol{z}_{\mathcal{G}}$ with respect to permutation and graph and signal manipulation

$$\boldsymbol{z}_{\mathcal{G}} = [\boldsymbol{z}^0 \oplus \cdots \oplus \boldsymbol{z}^L], \tag{4}$$

where $\oplus$ indicates concatenation and $\boldsymbol{z}^\ell$ is the $\ell-$th order coefficients, defined as

$$\boldsymbol{z}^\ell = [\mathcal{S}_{\ell,1} \oplus \cdots \oplus \mathcal{S}_{\ell,m_\ell}], \tag{5}$$

with $m_\ell = \mathcal{J}^\ell$ and $\mathcal{S}_{\ell,i}$ is defined as

$$\mathcal{S}_{\ell,i} = \eta \left(\rho \left(\dots \left(\rho \left(\boldsymbol{f} *_{\mathcal{G}} g_{j_1}\right) \dots\right) *_{\mathcal{G}} g_{j_\ell}\right)\right), \tag{6}$$

where $1 \leq i \leq m_\ell$. Here, L indicates the depth of the network, ρ is the modulus (absolute value) non-linearity, η stands for the average graph pooling operator, $1 \leq j_1, \dots j_\ell \leq \mathcal{J}$, and $\mathcal{S}_{0,1} = \eta(\boldsymbol{f})$.

A. Using the spectral filters $\{\hat{g}_j\}_{j=1}^{\mathcal{J}}$ defined in problem 1, construct scattering feature maps $\boldsymbol{z}_G$ from molecules in the QM9 dataset, using $\mathcal{J} = 8$ and $L = 2$.

B. Using PCA, map the feature maps $\boldsymbol{z}_{\mathcal{G}}$ onto a 2D plan $\boldsymbol{z}_{\mathcal{G}} \in \mathbb{R}^2$. Use scatter plot to visualize the 2D map of the scattering latent space by coloring each representation $\boldsymbol{z}_{\mathcal{G}}$ with the number of hydrogen-bond acceptor atoms corresponding to the molecular graph $\mathcal{G}$.

C. Develop a 2-layer neural network to predict the Highest Occupied Molecular Orbital (HOMO) energy of the molecule from scattering representations. Use 5000 data points to train the model. Plot the loss function.

D. Use 1000 test data points to evaluate the model. Show the results using a target (y-axis) vs. prediction (x-axis) scatter plot.