Machine Learning for Graphs and Sequential Data Exercise Sheet 6

Graphs: Embeddings and Classification

1 Node Embeddings

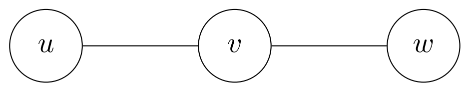

Figure 1: Undirected 3-chain for the Graph2Gauss problem

Problem 1: Consider an undirected 3-chain as in Figure 1 with three nodes u, v and w that we want to embed into $\mathbb{R}$, i.e. 1-dimensional, with Graph2Gauss. Find the embeddings analytically that we get by minimizing the training loss for a fixed embedding variance 1. So we are embedding each node as a 1-dimensional Gaussian with variance 1 by minimizing the loss

$$\mathcal{L} = E_{uv}^2 + e^{-E_{uw}} + E_{wv}^2 + e^{-E_{wu}}$$

where $E_{uv} = \mathrm{KL}(f(u)||f(v))$ is the KL divergence between the embeddings of node u and v.

Hint: The KL divergence between two normal distributions $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{N}(\nu, \tau^2)$ simplifies to

$$\mathrm{KL}\left(\mathcal{N}(\mu, \sigma^2)||\mathcal{N}(\nu, \tau^2)\right) = \log\frac{\tau}{\sigma} + \frac{\tau^2 + (\mu - \nu)^2}{2\sigma^2} - \frac{1}{2}.$$

Hint: Use the Lambert W-function to denote the inverse of $x\exp(x)$, i.e.

$$x\exp(x) = y \Rightarrow W(y) = x.$$

If you want to find a numerical solution, you can evaluate it for example on WolframAlpha with ProductLog(x).

Since the embedding variance is fixed, we only optimize over the means. Denote the embedding mean of node u by u and so on.

We begin by simplifying the KL divergence in this special case to

$$\mathrm{KL}(\mathcal{N}(u, 1)||\mathcal{N}(v, 1)) = \log 1 + \frac{1 + (u - v)^2}{2} - \frac{1}{2} = 0 + \frac{1}{2} + \frac{(u - v)^2}{2} - \frac{1}{2} = \frac{1}{2}(u - v)^2.$$

Plugging this into the loss $\mathcal{L}$ simplifies it to

$$\mathcal{L} = \frac{1}{4}(u - v)^4 + \exp\left(-\frac{(u - w)^2}{2}\right) + \frac{1}{4}(w - v)^4 + \exp\left(-\frac{(w - u)^2}{2}\right)$$

Collect common terms

$$= \frac{1}{4}(u - v)^4 + \frac{1}{4}(w - v)^4 + 2\exp\left(-\frac{(u - w)^2}{2}\right)$$

Only the fourth power terms depend on v, so it is easiest to minimize with respect to v first.

$$\frac{\partial}{\partial v}\left(\frac{1}{4}(u-v)^4 + \frac{1}{4}(w-v)^4\right) = -(u-v)^3 - (w-v)^3 = 0 \Leftrightarrow -(u-v) = w - v \Leftrightarrow v = \frac{u+w}{2}$$

Now that we have found the minimum of $\mathcal{L}$ in v as a function of u and w, we can reduce $\mathcal{L}$ to a two-dimensional problem.

$$\begin{aligned}
\mathcal{L} &= \frac{1}{4}\left(u - \frac{u+w}{2}\right)^4 + \frac{1}{4}\left(w - \frac{u+w}{2}\right)^4 + 2\exp\left(-\frac{1}{2}(u-w)^2\right) \\
&= \frac{1}{4}\left(\frac{u-w}{2}\right)^4 + \frac{1}{4}\left(\frac{w-u}{2}\right)^4 + 2\exp\left(-\frac{1}{2}(u-w)^2\right) \\
&= \frac{1}{2}\left(\frac{u-w}{2}\right)^4 + 2\exp\left(-\frac{1}{2}(u-w)^2\right) \\
&= \frac{1}{2^5}(u-w)^4 + 2\exp\left(-\frac{1}{2}(u-w)^2\right)
\end{aligned}$$

And we can even make it one-dimensional by reparameterizing with the difference $d = u - w$ between u and w.

$$\mathcal{L} = \frac{1}{2^5}d^4 + 2\exp\left(-\frac{d^2}{2}\right)$$

The first derivative of $\mathcal{L}$ in d is $\mathcal{L}' = 2^{-3}d^3 - 2d\exp\left(-\frac{d^2}{2}\right)$ which has a root at $d = 0$. By visualizing $\mathcal{L}$ as a parabola (fourth power) with a bump in the middle ($\exp(-d^2)$), we can eliminate $d = 0$ as a minimum. Assuming $d \neq 0$, we solve

$$2^{-3}d^3 - 2d\exp\left(-\frac{d^2}{2}\right) = 0 \Leftrightarrow \frac{1}{8} = \frac{2}{d^2}\exp\left(-\frac{d^2}{2}\right) \Leftrightarrow 8 = \frac{d^2}{2}\exp\left(\frac{d^2}{2}\right)$$

We denote the solution of this for $\frac{d^2}{2}$ by the Lambert W function with

$$\frac{d^2}{2} = W(8) \Rightarrow d = \sqrt{2W(8)}$$

where we have arbitrarily chosen the positive root since the embedding should be symmetric in u and w because they are at equivalent positions in the graph. So we finally arrive at the embeddings

$$u = w + d = w + \sqrt{2W(8)} \approx w + 1.7921$$

and

$$v = \frac{1}{2}(u+w) = w + \frac{1}{2}\sqrt{2W(8)} \approx w + 0.89605$$

with w as a free variable. As one would expect, Graph2Gauss embeds u and w symmetrically around v and w remains as a free variable because the loss only constrains differences between variables but does not impose any constraints regarding a reference point.

2 Label Propagation

Problem 2: The goal in Label Propagation is to find a labeling $\boldsymbol{y} \in \{0,1\}^N$ that minimizes the energy $\min_{\boldsymbol{y}} \frac{1}{2} \sum_{ij} \boldsymbol{w}_{ij}(y_i - y_j)^2$ subject to $y_i = \hat{y}_i \ \forall i \in S$ where the set of nodes V has been partitioned into the labeled nodes S and the unlabeled nodes U, $w_{ij} \geq 0$ is the non-negative edge weight and $\hat{y}_i$ are the observed labels.

Following from the first observation regarding the Laplacian, the minimization problem can be rewritten and then relaxed to $\min_{\boldsymbol{y} \in \mathbb{R}^N} \boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y}$ subject to the same constraints. Show that the closed form solution is

$$\boldsymbol{y}_U = -\boldsymbol{L}_{UU}^{-1} \cdot \boldsymbol{L}_{US} \cdot \hat{\boldsymbol{y}}_S$$

where w.l.o.g. we assume that the Laplacian matrix is partitioned into blocks for labeled and unlabeled nodes as

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}_{SS} & \boldsymbol{L}_{SU} \\ \boldsymbol{L}_{US} & \boldsymbol{L}_{UU} \end{pmatrix}.$$

We begin by plugging the block partitioned form of $\boldsymbol{L}$ into the minimization term.

$$\begin{aligned} \boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y} &= \boldsymbol{y}^T \begin{pmatrix} \boldsymbol{L}_{SS} & \boldsymbol{L}_{SU} \\ \boldsymbol{L}_{US} & \boldsymbol{L}_{UU} \end{pmatrix} \boldsymbol{y} \\ &= \hat{\boldsymbol{y}}_S^T \boldsymbol{L}_{SS} \hat{\boldsymbol{y}}_S + \hat{\boldsymbol{y}}_S^T \boldsymbol{L}_{SU} \boldsymbol{y}_U + \boldsymbol{y}_U^T \boldsymbol{L}_{US} \hat{\boldsymbol{y}}_S + \boldsymbol{y}_U^{\mathrm{T}} \boldsymbol{L}_{UU} \boldsymbol{y}_U \end{aligned}$$

The laplacian is symmetric and therefore $\boldsymbol{L}_{US} = \boldsymbol{L}_{SU}^T$.

$$= \hat{\boldsymbol{y}}_S^T \boldsymbol{L}_{SS} \hat{\boldsymbol{y}}_S + 2\boldsymbol{y}_U^T \boldsymbol{L}_{US} \hat{\boldsymbol{y}}_S + \boldsymbol{y}_U^{\mathrm{T}} \boldsymbol{L}_{UU} \boldsymbol{y}_U =: f(\boldsymbol{y}_U)$$

We can find the minimizer of f by finding the root of its first derivative with respect to $\boldsymbol{y}_U$ because f is quadratic.

$$\frac{\partial f}{\partial \boldsymbol{y}_U} = 2\boldsymbol{L}_{US} \hat{\boldsymbol{y}}_S + \left(\boldsymbol{L}_{UU} + \boldsymbol{L}_{UU}^T\right) \boldsymbol{y}_U = 2\boldsymbol{L}_{US} \hat{\boldsymbol{y}}_S + 2\boldsymbol{L}_{UU} \boldsymbol{y}_U = 0 \Leftrightarrow \boldsymbol{y}_U = -\boldsymbol{L}_{UU}^{-1} \cdot \boldsymbol{L}_{US} \cdot \hat{\boldsymbol{y}}_S$$

3 Spectral GNNs

Problem 3: Consider the spectral GNN given by

$$\boldsymbol{Z} = \phi(\boldsymbol{U} g(\boldsymbol{\Lambda}) \boldsymbol{U}^T \varphi(\boldsymbol{X})),$$

where ϕ and φ are non-linear, parametrized functions, e.g. multi-layer perceptrons. For this exercise we choose a polynomial filter of the form

$$g(\lambda) = \sum_{k=0}^{\infty} \theta_k \lambda^k.$$

Note that instead of parametrizing the spectral filter g we can also choose fixed coefficients θ_k, for example

$$\theta_k = \frac{(-t)^k}{k!}$$

where $t > 0$ is a hyperparameter that we can fine-tune.

Show that this choice of g constraints the possible graph filters.

We first observe that for the spectral filter g we have

$$g(\lambda) = \sum_{k=0}^{\infty} \theta_k \lambda^k = \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} \lambda^k = \sum_{k=0}^{\infty} \frac{(-t\lambda)^k}{k!} = e^{-t\lambda}$$

Note that for all $\lambda_i < \lambda_j$ we have $g(\lambda_i) > g(\lambda_j)$ since

$$\frac{g(\lambda_i)}{g(\lambda_j)} = e^{t(\lambda_j - \lambda_i)} > 1$$

(Note that the fraction is well-defined due to the definition of the exponential function.)

Since $g(\lambda_i) > g(\lambda_j)$ for any $\lambda_i < \lambda_j$, g corresponds to a low-pass filter and is therefore constrained. The larger the hyperparameter t the more we will diminish large eigenvalues.

4 PPNP

Problem 4: The iterative equation of PPNP is given by

$$\boldsymbol{H}^{(l+1)} = (1-\alpha)\hat{\boldsymbol{A}}\boldsymbol{H}^{(l)} + \alpha \boldsymbol{H}^{(0)}$$

where $\hat{\boldsymbol{A}} = \tilde{\boldsymbol{D}}^{-\frac{1}{2}} \tilde{\boldsymbol{A}} \tilde{\boldsymbol{D}}^{-\frac{1}{2}}$ is the propagation matrix. Derive the closed form solution for infinitely many propagation steps.

Hint: If we have for a matrix $\boldsymbol{T}$ that all its eigenvalues λ are strictly between -1 and 1, an equivalent matrix formulation of the geometric series formula holds and

$$\sum_{k=0}^{\infty} \boldsymbol{T}^k = (\boldsymbol{I} - \boldsymbol{T})^{-1}.$$

Hint: The eigenvalues λ_i of any normalized Laplacian $\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{A} \boldsymbol{D}^{-\frac{1}{2}}$ are $0 \le \lambda_i \le 2$.

We start with $H^{(1)}$ and expand for a few steps.

$$\boldsymbol{H}^{(1)} = (1-\alpha)\hat{\boldsymbol{A}}\boldsymbol{H}^{(0)} + \alpha\boldsymbol{H}^{(0)}$$

$$\begin{aligned}
\boldsymbol{H}^{(2)} &= (1-\alpha)\hat{\boldsymbol{A}}\boldsymbol{H}^{(1)} + \alpha\boldsymbol{H}^{(0)} \\
&= (1-\alpha)\hat{\boldsymbol{A}}\left((1-\alpha)\hat{\boldsymbol{A}}\boldsymbol{H}^{(0)} + \alpha H^{(0)}\right) + \alpha\boldsymbol{H}^{(0)} \\
&= (1-\alpha)^2\hat{\boldsymbol{A}}^2\boldsymbol{H}^{(0)} + (1-\alpha)\hat{\boldsymbol{A}}\alpha\boldsymbol{H}^{(0)} + \alpha\boldsymbol{H}^{(0)}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{H}^{(3)} &= (1-\alpha)\hat{\boldsymbol{A}}\boldsymbol{H}^{(2)} + \alpha\boldsymbol{H}^{(0)} \\
&= (1-\alpha)\hat{\boldsymbol{A}}\left((1-\alpha)^2\hat{\boldsymbol{A}}^2\boldsymbol{H}^{(0)} + (1-\alpha)\hat{\boldsymbol{A}}\alpha\boldsymbol{H}^{(0)} + \alpha\boldsymbol{H}^{(0)}\right) + \alpha\boldsymbol{H}^{(0)} \\
&= (1-\alpha)^3\hat{\boldsymbol{A}}^3\boldsymbol{H}^{(0)} + (1-\alpha)^2\hat{\boldsymbol{A}}^2\alpha\boldsymbol{H}^{(0)} + (1-\alpha)\hat{\boldsymbol{A}}\alpha\boldsymbol{H}^{(0)} + \alpha\boldsymbol{H}^{(0)}
\end{aligned}$$

We can see the following pattern emerge.

$$\boldsymbol{H}^{(k)} = \left((1-\alpha)\hat{\boldsymbol{A}}\right)^k\boldsymbol{H}^{(0)} + \left(\sum_{i=0}^{k-1}\left((1-\alpha)\hat{\boldsymbol{A}}\right)^i\right)\alpha\boldsymbol{H}^{(0)}$$

If we let k grow to infinity, the first term converges to 0 because $\alpha \in (0,1)$ and in the second term we can apply the geometric series formula to get

$$\boldsymbol{H}^{(\infty)} = \alpha\left(\boldsymbol{I} - (1-\alpha)\hat{\boldsymbol{A}}\right)^{-1}\boldsymbol{H}^{(0)}$$

as the closed form solution as long as the eigenvalues of $(1-\alpha)\hat{\boldsymbol{A}}$ are strictly between -1 and 1.

We know from the hint that the eigenvalues of the normalized Laplacian $\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}$ for any graph structure A are in $[0,2]$. So it is also true for the amended graph structure with self-loops $\hat{\boldsymbol{A}}$ where $\boldsymbol{L} = \boldsymbol{I} - \hat{\boldsymbol{A}}$. Let λ be an eigenvalue of $\hat{\boldsymbol{A}}$ with eigenvector $\boldsymbol{v}$.

$$(1-\lambda)\boldsymbol{v} = \boldsymbol{v} - \hat{\boldsymbol{A}}\boldsymbol{v} = \left(\boldsymbol{I} - \hat{\boldsymbol{A}}\right)\boldsymbol{v} = \boldsymbol{L}\boldsymbol{v}$$

So $1-\lambda$ is also an eigenvector $\boldsymbol{L}$ and must therefore be in $[0,2]$. Consequently, the eigenvalues of $\hat{\boldsymbol{A}}$ are in $[-1,1]$ and the eigenvalues of $(1-\alpha)\hat{\boldsymbol{A}}$ are in $(-1,1)$ because $0 < (1-\alpha) < 1$.