Machine Learning for Graphs and Sequential Data Exercise Sheet 6

Graphs: Embeddings and Classification

---

1 Node Embeddings

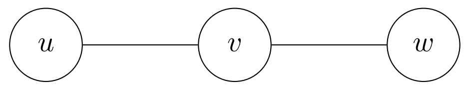


Figure 1: Undirected 3-chain for the Graph2Gauss problem

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).

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

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

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.

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 $0 \leq \lambda_i \leq 2$.

1 Node Embeddings

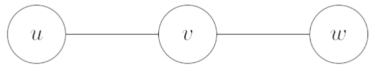


Figure 1: Undirected 3-chain for the Graph2Gauss problem

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

$$\boxed{\mathcal{L}} = E_{uv}^2 + e^{-E_{uw}} + E_{wv}^2 + e^{-E_{wu}}$$ no s.t

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).

$$KL(N(\mu,1)||N(v,1)) = \frac{(\mu-v)^2}{2}$$

$$L = \frac{1}{4}(\mu-v)^4 + e^{-\frac{(\mu-w)^2}{2}} + \frac{1}{4}(w-v)^4 + e^{-\frac{(w-u)^2}{2}}$$

$$= \frac{1}{4}(\mu-v)^4 + \frac{1}{4}(w-v)^4 + 2e^{-\frac{(\mu-w)^2}{2}}$$

$$\frac{\partial L}{\partial v} = -(\mu-v)^3 - (w-v)^3 \doteq 0$$

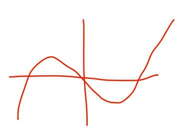
$$-(\mu-v) = w-v$$

$$v = \frac{u+w}{2}$$

$$L = \frac{1}{4}\left(\frac{\mu-w}{2}\right)^4 + \frac{1}{4}\left(\frac{w-\mu}{2}\right)^4 + 2e^{-\frac{(\mu-w)^2}{2}}$$

$$= \frac{1}{2^5}(\mu-w)^4 + 2e^{-\frac{(\mu-w)^2}{2}} \Leftarrow d = \mu-w$$

$$= \frac{1}{2^5}d^4 + 2e^{-\frac{d^2}{2}} \Leftarrow$$



min is not 0

$$\frac{\partial L}{\partial d} = \frac{1}{2^3}d^3 + 2e^{-\frac{1}{2}d^2} \cdot -d \doteq 0$$

$d \neq 0$

$$\frac{1}{8}d^3 = 2d\,e^{-\frac{1}{2}d^2}$$

$$\frac{1}{8} = \frac{2}{d^2}e^{-\frac{d^2}{2}}$$

LW $\Rightarrow$

$$8 = \frac{d^2}{2}e^{\frac{d^2}{2}}$$

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

$$\begin{cases} \mu - w = d = \sqrt{2W(8)} \\ v = \frac{\mu+w}{2} \end{cases}$$

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

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

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}_{SS} & \boldsymbol{L}_{SU} \\ \boldsymbol{L}_{US} & \boldsymbol{L}_{UU} \end{pmatrix}. \quad y = \begin{bmatrix} \hat{y}_s \\ y_u \end{bmatrix}$$

$y = 1$

$y = 2$

$$\min_{y \in \mathbb{R}^N} y^T L y = \min_{y_u} [\hat{y}_s \;\; y_u] \begin{bmatrix} L_{ss} & L_{su} \\ L_{us} & L_{uu} \end{bmatrix} \begin{bmatrix} \hat{y}_s \\ y_u \end{bmatrix}$$

$$= \min_{y_u} [L_{ss}\hat{y}_s + L_{us} y_u \quad L_{su}\hat{y}_s + L_{uu} y_u] \begin{bmatrix} \hat{y}_s \\ y_u \end{bmatrix}$$

$$= \min_{y_u} L_{ss}\hat{y}_s^2 + L_{us} y_u \cdot \hat{y}_s + L_{su}\hat{y}_s y_u + L_{uu} y_u^2$$

$$\frac{\partial \cdots}{\partial y_u} = L_{us}\hat{y}_s + L_{su}\hat{y}_s + 2L_{uu} y_u \overset{!}{=} 0$$

$$= 2L_{us}\hat{y}_s + 2L_{uu} y_u \overset{!}{=} 0.$$

$$y_u = -L_{uu}^{-1} L_{us} y_u$$

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

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

tylor

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

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

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

For all $\lambda_i < \lambda_j$, $g(\alpha_i) > g(\lambda_j)$

PPNP

Problem 3: 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 $0 \leq \lambda_i \leq 2$.

$$H^0 = \phi(X)$$

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

$$H^2 = (1-\alpha)\hat{A}H^1 + \alpha H^0$$

$$= (1-\alpha)\hat{A}\left[(1-\alpha)\hat{A}H^0 + \alpha H^0\right] + \alpha H^0$$

$$= (1-\alpha)^2\hat{A}\hat{A}H^0 + (1-\alpha)\alpha\hat{A}H^0 + \alpha H^0$$

$$H^n = \underbrace{(1-\alpha)^n}_{\downarrow\, 0}\hat{A}^n H^0 + \underbrace{\left(\alpha\sum_{j=1}^{n-1}(1-\alpha)^j\hat{A}^j\right)H^0 + \alpha H^0}_{\alpha\left[I-(1-\alpha)\hat{A}\right]^{-1}H^0}$$

$n \to \infty$

$$T = (1-\alpha)\hat{A}$$

$$\hat{A}v = \lambda v$$

$$L = I - \hat{A}$$

$$Lv = (I-\hat{A})v = v - \hat{A}v = v - \lambda v = \underbrace{(1-\lambda)}_{\lambda'}v$$

$$0 \leq \lambda' \leq 2$$

$$-1 \leq \lambda \leq 1$$

eigenvalues of $\hat{A} \in [-1,1]$

$\alpha \in (0,1)$

$1-\alpha \in (0,1)$

eigenvalues of $(1-\alpha)\hat{A} \in (-1,1)$