

# 机器学习第六次作业

郭洲蕊 2021E8014682024

一、四个图卷积网络解决的问题，训练数据、优化目标

1、顶点分类：利用图的结构和顶点的先验信息(对应于一个多维向量)去对每个顶点进行分类。

训练数据：图近邻结构、顶点的先验信息

优化目标：最小化顶点分类损失，例如交叉熵

2、链路预测：利用图的相邻结构或者图随着时间结构演化来预测两个顶点间存在边的可能性

训练数据：部分已知图的近邻结构、顶点的先验信息

优化目标：最小化预测边的错误率，例如最小化两类别交叉熵（存在/不存在边）

3、社区划分：利用图的相邻结构对整个图划分成多个子图使得子图内的顶点相互关系紧密，子图间的顶点的关系稀疏

训练数据：图的近邻结构

优化目标：最小化跨越不同社区的边的数目

4、图匹配：给定两个或多个图，利用每幅图的结果和顶点的先验信息去寻找不同图上的顶点之间的匹配关系

训练数据：图近邻结构、顶点的先验信息

优化目标：最小化匹配错误的顶点对数量

二、N个点的邻接矩阵A 嵌入矩阵Z 证 $Z^T Z$ 是A的低阶秩近似

损失函数： $L = ||Z^T Z - A||_F^2$

损失函数最小的时候Z有最优解，满足低阶秩近似的定义，所以 $Z^T Z$ 是A的低阶秩近似。

三、图神经网络消息传递公式，不同方法的不同点

$$h_v^0 = [F]_v$$

$$h_v^k = \sigma \left( W_k \sum_{u \in N(v)} \frac{h_u^{k-1}}{|N(v)|} + B_k h_v^{k-1} \right), k = 1, 2, \dots, K$$

$$z_k = h_v^K$$

F为顶点原始信息向量，是输入变量

v是顶点，h是本层网络输出，k代表第几层网络

W和B是可学习参数

N(v)是v点的邻居顶点集合， $\sigma$ 是任意非线性变换

Z是输出的顶点的特征向量

基础方法、Kipfs 方法、GraphSAGE 方法的消息传递公式不同：

基础方法：

$$\mathbf{h}_v^k = \sigma \left( \mathbf{W}_k \sum_{u \in N(v)} \frac{\mathbf{h}_u^{k-1}}{|N(v)|} + \mathbf{B}_k \mathbf{h}_v^{k-1} \right)$$

Kipfs 方法：

$$\mathbf{h}_v^k = \sigma \left( \mathbf{W}_k \sum_{u \in N(v) \cup v} \frac{\mathbf{h}_u^{k-1}}{\sqrt{|N(u)||N(v)|}} \right)$$

GraphSAGE：

$$\mathbf{h}_v^k = \sigma \left( [\mathbf{W}_k \cdot \text{AGG}(\{\mathbf{h}_u^{k-1}, \forall u \in N(v)\}), \mathbf{B}_k \mathbf{h}_v^{k-1}] \right)$$

Kipfs 方法比基础方法少学习参数 B，GraphSAGE 方法把邻点和自身的影响单独算，基础公式是一起加权起来的

四、3 种自编码器对隐变量的约束，损失函数

1、模型约束：压缩编码

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2 &= \|\mathbf{X} - \mathbf{W}^\top \mathbf{W} \mathbf{X}\|^2 \\ &= \|\mathbf{X} \mathbf{X}^\top - \mathbf{W}^\top \mathbf{W}\|^2 + C \end{aligned}$$

2、正则项约束：稀疏编码器、隐空间平滑约束

$$L(\mathbf{x}, g(f(\mathbf{x}))) + \Omega(\mathbf{h})$$

$$\Omega(\mathbf{h}) = \lambda \sum_i |h_i|$$

3、数据、训练任务约束：去噪编码器

---


$$L(\mathbf{x}, g(f(\tilde{\mathbf{x}})))$$

$$\tilde{\mathbf{x}} \sim C(\tilde{\mathbf{x}}|\mathbf{x})$$

五、 推导变分自编码器的损失函数

定义证据下界：

$$L(D, E|\mathbf{x}) = \mathbb{E}_{\mathbf{h} \sim E(\mathbf{h}|\mathbf{x})} [-\log(E(\mathbf{h}|\mathbf{x})) + \log D(\mathbf{h}, \mathbf{x})]$$

因为 KL 散度非负：

$$KL[E(\mathbf{h}|\mathbf{x})||D(\mathbf{h}|\mathbf{x})] = \mathbb{E}_{\mathbf{h} \sim E(\mathbf{h}|\mathbf{x})} \log \frac{E(\mathbf{h}|\mathbf{x})}{D(\mathbf{h}|\mathbf{x})} \geq 0$$

又因为  $D(\mathbf{h}|\mathbf{x}) = D(\mathbf{h}, \mathbf{x})/D(\mathbf{x})$ ，

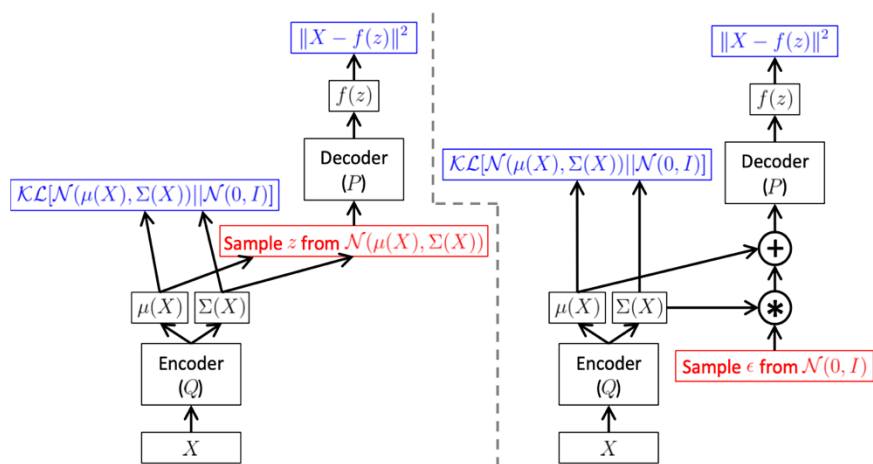
$$\begin{aligned} KL[E(\mathbf{h}|\mathbf{x})|D(\mathbf{h}|\mathbf{x})] &= \sum_{\mathbf{h}} E(\mathbf{h}|\mathbf{x}) \log \frac{E(\mathbf{h}|\mathbf{x})}{D(\mathbf{h}|\mathbf{x})} \\ &= \sum_{\mathbf{h}} E(\mathbf{h}|\mathbf{x}) \log \frac{E(\mathbf{h}|\mathbf{x})D(\mathbf{x})}{D(\mathbf{h}, \mathbf{x})D(\mathbf{x})} \\ &= \sum_{\mathbf{h}} E(\mathbf{h}|\mathbf{x}) \log \frac{E(\mathbf{h}|\mathbf{x})}{D(\mathbf{h}, \mathbf{x})} + \sum_{\mathbf{h}} E(\mathbf{h}|\mathbf{x}) \log D(\mathbf{x}) \\ &= -L(E, D|\mathbf{x}) + \log D(\mathbf{x}) \end{aligned}$$

所以：

$$\log D(\mathbf{x}) = KL[E(\mathbf{h}|\mathbf{x})||D(\mathbf{h}|\mathbf{x})] + L(D, E|\mathbf{x})$$

即：

$$\log D(\mathbf{x}) \geq L(D, E|\mathbf{x})$$



左图红色部分对隐变量  $z$  进行抽样  $z \sim \mathcal{N}(\mu(X), \Sigma(X))$ ，十分随机，会造成梯度反传中断，无法直接端到端训练

$$z = \mu(X) + \Sigma^{\frac{1}{2}}(X)\epsilon$$

右图重抽样方法  $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ ，可以保持梯度信息的连续

## 六、 GAN 的损失函数

$$\min_{\tilde{\mathbf{w}}} \max_{\mathbf{w}} \mathbb{E}_{\mathbf{y} \sim P_{\text{real}}} F(D(\mathbf{y}; \mathbf{w})) - \mathbb{E}_{\boldsymbol{\eta} \sim P_{\text{simple-distribution}}} \tilde{F}(D(G(\boldsymbol{\eta}; \tilde{\mathbf{w}}); \mathbf{w}))$$

D 是判别器，参数是  $\mathbf{w}$

G 是生成器，参数是  $\tilde{\mathbf{w}}$

$\mathbf{y}$  是真实样本， $\boldsymbol{\eta}$  是生成器的初始值

F 和  $\tilde{F}$  是表示判断结果的价值权重

训练问题：

- 1、梯度消失
- 2、震荡、不收敛
- 3、模式坍塌
- 4、高分辨率模型训练速度慢