机器学习 第七次作业

黄磊 计702 2022E8013282156

题目 1. 列举四个图卷积神经网络解决的问题，并简述这些问题相应的训练数据和优化目标是什么样的。

问题：整图级别的问题，例如图的分类；顶点级别的问题，例如顶点分类、社群划分；边级别的问题，例如链路预测；图上的随机优化问题，例如图匹配、最优路径；图上的动态预测问题，例如SIR传染病模型。

对于顶点级别中的顶点分类任务：训练数据是图近邻结构和顶点的先验知识，优化目标是最小化顶点分类损失，例如交叉熵；

对于顶点级别中的社区划分任务：训练数据是图的近邻结构，优化目标是最小化跨越不同社区边的数目；

对于边层次中的链路预测任务：训练数据为部分已知图的近邻结构和顶点的先验知识，优化目标为最小化预测边的错误率；

对于图模型组合优化问题中的图匹配，训练数据为图近邻结构和顶点的先验知识，优化目标为最小化匹配错误的顶点对数量；

对于传染病传递模型，训练数据为图结构和节点的历史数据，优化目标为预测未来一段时间节点的状态信息。

题目 2. 已知一个有 N 个顶点的图的邻接矩阵为 A，设基于邻接矩阵的图嵌入给出的嵌入矩阵为 $Z \in R^{P\times N}$ 。证明Z^TZ是矩阵 A 的低阶秩近似。

A是一个邻接矩阵，因此是对称的，将A进行特征值分解，得到：

$$A = U\Lambda U^T$$

其中，U是正交矩阵，Λ 是对角矩阵，对角线元素为A的特征值。

再考虑矩阵 Z^TZ ，其第 i 行第 j 列的元素是矩阵Z中第 i 个顶点和第 j 个顶点的相似度加权和，其中权重是嵌入矩阵的每一行的元素：

$$(Z^TZ)_{ij} = \sum_{k=1}^{P} z_{ki}z_{kj}$$

将其分解为：

$$(Z^TZ) = (Z^TU)(U^TZ)$$

U^T Z 是一个 N×P 的矩阵，其第 i 行第 j 列的元素为 U 的第 i 行和 z_j 的第 j 列的内积。因为 U 是正交矩阵，所以 $U^T U = I$，因此 U^T Z 的第 i 行第 j 列的元素可以表示为 U 的第 i 行和第 j 个顶点的嵌入向量的内积。

接下来考虑 Z^TZ 的秩，其中Z为 $P \times N$ 的矩阵，因此秩不会超过 P ，同时，如果 A 中有 k 个0特征值，则 A 的秩不会超过 N−k ，因此如果选择 $P = N - k$ ，则 Z^TZ 的秩不会超过 $N - k$。

考虑 Z^TZ 与 A 之间的关系，可以把Z^TZ表示为：

$$Z^TZ = (Z^TU)(U^TZ) = V\Lambda V^T$$

其中，$V = Z^T U$ 是一个 P×N 的矩阵，Λ 是一个对角矩阵，其对角线上的元素为 $Z^T Z$ 的特征值。因为 $Z^T Z$ 是一个 P×P 的矩阵，所以它的特征值和特征向量可以直接计算。因为 $Z^T Z$ 是矩阵 A 的低阶秩近似，所以它的特征值中只有少数几个非零的特征值。因此，我们可以选择 P=N-k，使得 $Z^T Z$ 的秩不会超过 N-k。因为 A 的秩也不会超过 N-k，所以我们可以认为 $Z^T Z$ 是矩阵 A 的低阶秩近似。

题目 3. 写出图神经网络消息传递公式，并指出公式中每个变量的含义，以及哪些变量是可学习变量，哪些变量是输入变量；指出基础的图神经网络消息传递方法，Kipf 方法，GraphSAGE方法的不同点。

题目 4：列举 3 种自编码器中对隐变量的约束，写出它们对应的损失函数。

1. 稀疏性约束：让隐变量大多数数值为0得到更为紧凑的表示。损失函数加入稀疏惩罚项即可，

例如 L_1 正则项，h为隐变量：

$$L = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x_i}|^2 + \lambda|h|_1$$

2. 降噪约束：在自编码器中加入一些随机性，使得模型对噪声更加鲁棒。通过在输入数据中加入噪声并在输出时候重构无噪声的输入，损失函数为：

$$L = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x_i}|^2$$

3. 边缘分布约束：指限制隐变量的边缘分布，使得隐变量能够学习到数据的主要因素。边缘分布约束可以通过在损失函数中加入一个最大化边缘似然的项来实现，例如：

$$L = -logp(\boldsymbol{x}) + \lambda D_{KL}(q(\boldsymbol{h} \mid \boldsymbol{x}) \,||\, p(\boldsymbol{h}))$$

其中 p (x) 是输入数据的边缘分布， q(h|x) 是给定输入数据 x 后隐变量的后验分布， p(h) 是隐变量的先验分布， D_{KL} 是 KL 散度，λ 是控制正则化的超参数。这个损失函数包含了两个部分，第一个部分是最小化重构误差，第二个部分是最大化隐变量的独立性，从而提高模型的泛化能力。

题目 5：假设 VAE 编码器在给定输入x其隐变量h的分布为$E(h|x)$，解码器在给定隐变量h其对应的x的分布为$D(x|h)$，模型的隐变量满足$Q(h)$。设模型后验概率为$D(x) = \int D(x|h)Q(h)dh$ 。请从“证据下界”角度推导变分自编码器的损失函数。

通过最大化ELBO来间接实现最大化数据的对数似然。根据贝叶斯公式：

$$D(h|x) = D(x|h)D(h)/D(x)$$

$ELBO$ 表示为：

$$ELBO = E[\log D(x|h)] - KL[Q(h|x)||D(h|x)]$$

由于 KL 散度始终为负值，因此：

$$\log P(x) \geq ELBO$$

因此最大化 $\log P(x)$ 可以通过最大化 ELBO 来实现，根据贝叶斯重写ELBO：

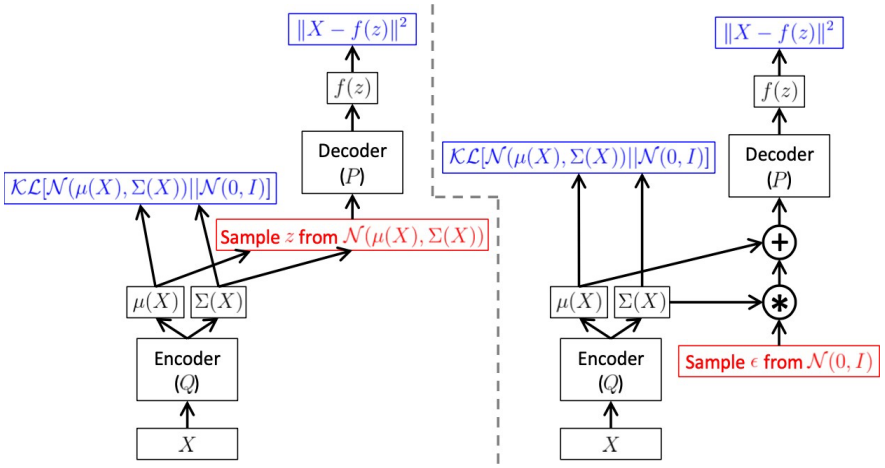
$$ELBO = E[\log D(x \mid h)] - KL[(Q(h \mid x) \dashv |D(h)D(x \mid h))/ D(x)]$$
$$= \mathrm{E}[\log D(\mathrm{x}|\mathrm{h})] - \mathrm{KL}[\mathrm{Q}(\mathrm{h}|\mathrm{x}) \,||\, D(\mathrm{h})] + \mathrm{KL}[\mathrm{Q}(\mathrm{h}|\mathrm{x}) \,||\, D(\mathrm{x}|\mathrm{h})]$$

第三项与 h 无关，因此我们可以只关注前两项，据此给出损失函数：

$$L(x) = -E[\log D(x|h)] + KL[Q(h|x) \,||\, D(h)]$$

其第一项是重构损失，第二项看作正则项。

下图是在实现 VAE 时常用的重抽样变换的示意图，解释 VAE 训练为什么需要做这样的变换，具体做了什么样的改变？



该变换是从潜在空间分布中采样生成随机噪声，并输入解码器中。通过重参数化技巧，我们可以将从潜在空间中采样的随机变量转换为对其参数的函数的采样，这样就可以应用基于梯度的优化算法来训练模型。而如果我们直接从潜在空间中进行采样，则无法使用基于梯度的优化算法来更新参数。

左图从潜在空间中抽样出噪声，右图则是从标准正态分布中抽样出噪声，并与潜在空间分布叠加。

题目 6：写出生成生成对抗模型（GAN）的损失函数，指明每个变量的意义。列出训练 GAN 时常出现的问题

生成对抗模型（GAN）的损失函数由两个部分组成：生成器损失和判别器损失。其中，生成器损失用于衡量生成器生成的样本与真实样本之间的距离，而判别器损失用于衡量判别器将生成样本与真实样本区分开的能力。

生成器损失函数的形式为：

$$LG = -\frac{1}{m}\sum_{i=1}^{m}\log D\left(G\left(z^{(i)}\right)\right)$$

其中，m 表示训练批次的大小，$z^{(i)}$ 表示从潜在空间中采样的第 i 个噪声向量，$G\left(z^{(i)}\right)$ 表示生成器生成的第 i 个样本，D 表示判别器。生成器的目标是最小化损失函数，从而生成更接近真实样本的样本。

判别器损失的形式为：

$$LD = -\frac{1}{m}\sum_{i=1}^{m}\left[D\left(x^{(i)}\right) + \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right)\right]$$

其中，$x^{(i)}$表示第 i 个真实样本，$D(x^{(i)})$ 表示判别器给出真实样本的概率，$1 - D\left(G\left(z^{(i)}\right)\right)$表示判别器给出生成样本的概率。判别器的目标是最大化 LD，从而更好地区分真实样本和生成样本。

在训练GAN时，常见的问题包括：

模式崩溃（mode collapse）：生成器只能生成有限种类的样本，而不能生成多样化的样本。
梯度消失和梯度爆炸：训练GAN时，判别器和生成器的梯度可能会变得非常小或非常大，从而导致训练不稳定。
对抗样本攻击：由于生成器和判别器都是基于神经网络的，因此它们容易受到对抗样本攻击，从而产生偏差或降低性能。
训练时间长：GAN需要大量的训练时间来达到良好的性能，这使得它们难以在实践中应用。