

C15+C16 机器学习作业

黄磊 计702 2022E8013282156

题目 1: 下图是无监督判别式学习 SimCLR 模型的示意图。依据示意图, 写出SimCLR 的损失函数并简述 SimCLR 模型是如何实现、训练的。

参考文献: Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey Hinton, A Simple Framework for Contrastive Learning of Visual Representations, ICML, 2020. <http://proceedings.mlr.press/v119/chen20j/chen20j.pdf>

SimCLR 对输入数据同时送入两个同源的数据扩增模块得到两种不同视角下的数据, 然后通过相同的神经网络模型编码得到特征向量, 再送入一个映射层将特征映射到隐空间, 得到投影向量 z_i 和 z_j 。实验证明, 在隐空间计算对比损失比在编码器后计算效果更好。

定义相似度计算方式:

$$\text{sim}(\mathbf{u}, \mathbf{v}) = \frac{(\mathbf{u}^T \mathbf{v})}{(|\mathbf{u}|) \cdot (|\mathbf{v}|)}$$

其损失函数为:

$$l(i, j) = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbf{1}_{k \neq i} \exp(\text{sim}(z_i, z_k)/\tau)}$$

SimCLR 模型的训练是通过交替地最小化对比损失和正则化损失来进行的。具体地, 对于每个批次的图像, 我们首先通过数据增强得到两个不同视角的图像, 然后将它们分别通过相同的神经网络模型得到其特征向量, 再将这些特征向量通过投影头投影到低维特征空间中。然后, 我们使用对比损失来测量同一图像的不同视角之间的相似度, 并通过反向传播更新模型参数来最小化损失。同时, 我们还使用正则化损失来约束特征向量的范数, 以防止模型出现过拟合的情况。

整个训练过程包括多个训练阶段, 每个阶段的训练数据不同, 且训练数据会逐渐变得更加难以区分。这种训练方式称为“自适应数据增强”, 其目的是让模型逐渐学习到更加抽象和通用的特征, 从而提高模型的泛化性能。最终, 我们可以使用训练好的模型来提取图像的特征向量, 并将其用于各种图像相关任务, 如图像分类、目标检测、图像分割等。

题目2:玻尔兹曼分布表示为 $P(\mathbf{x}) = \exp(-\beta H(\mathbf{x}))/Z$, 假设该玻尔兹曼分布的能量函数为 $H(\mathbf{x}) = -\sum_{i < j < k} J_{i,j,k} x_i x_j x_k$.

(1) 根据指数族(exponential family)定义,指出该模型的自然参数与充分统计量分别是什么?

指数分布族定义为:

$$P(\mathbf{x}; \boldsymbol{\eta}) = b(\mathbf{x}) e^{\boldsymbol{\eta}^T T(\mathbf{x}) - \alpha(\boldsymbol{\eta})}$$

对玻尔兹曼分布变形得到:

$$P(\mathbf{x}) = \exp(-\beta H(\mathbf{x}) - (-\ln Z))$$

因此, 其自然参数为: $\boldsymbol{\eta} = -\beta$, 充分统计量为 $H(\mathbf{x})$, 对数配分函数为 $A(\theta) = -\ln Z$.

(2) 推导 $\frac{\partial \log Z}{\partial J_{i,j,k}}$, 使用统计量 $\langle x_i, x_j, x_k \rangle$ 表示, 其中: $\langle x_i, x_j, x_k \rangle = \sum_{\mathbf{x}} P(\mathbf{x}) x_i x_j x_k$.

玻尔兹曼分布的配分函数 Z 可以写成:

$$Z = \sum_{\mathbf{x}} e^{-\beta H(\mathbf{x})}$$

其中, $H(\mathbf{x}) = -\sum_{i < j < k} J_{i,j,k} x_i x_j x_k$ 是能量函数, $x_i \in \{-1, 1\}$ 是状态 \mathbf{x} 的第 i 个分量, $\beta = \frac{1}{k_B T}$ 是反比于温度的常数, k_B 是玻尔兹曼常数。

我们可以对 $\log Z$ 关于 $J_{i,j,k}$ 求偏导数, 得到:

$$\frac{\partial \log Z}{\partial J_{i,j,k}} = \frac{\partial}{\partial J_{i,j,k}} \log \sum_{\mathbf{x}} e^{-\beta H(\mathbf{x})}$$

由于 \log 是一个单调递增函数, 它不会改变函数的最大值或最小值, 因此, 我们可以将 \log 作用于 $\sum_{\mathbf{x}} e^{-\beta H(\mathbf{x})}$, 得到:

$$\frac{\partial \log Z}{\partial J_{i,j,k}} = \frac{1}{\sum_{\mathbf{x}} e^{-\beta H(\mathbf{x})}} \frac{\partial}{\partial J_{i,j,k}} \sum_{\mathbf{x}} e^{-\beta H(\mathbf{x})}$$

注意到 $\frac{\partial H(\mathbf{x})}{\partial J_{i,j,k}} = -x_i x_j x_k$, 因此:

$$\frac{\partial \log Z}{\partial J_{i,j,k}} = \frac{1}{\sum_{\mathbf{x}} e^{-\beta H(\mathbf{x})}} \sum_{\mathbf{x}} \frac{\partial}{\partial J_{i,j,k}} e^{-\beta H(\mathbf{x})} = \frac{1}{\sum_{\mathbf{x}} e^{-\beta H(\mathbf{x})}} \sum_{\mathbf{x}} (-\beta x_i x_j x_k e^{-\beta H(\mathbf{x})})$$

将能量函数 $H(\mathbf{x})$ 代入上式可得:

$$\frac{\partial \log Z}{\partial J_{i,j,k}} = -\beta \frac{\sum_{\mathbf{x}} x_i x_j x_k e^{-\beta H(\mathbf{x})}}{\sum_{\mathbf{x}} e^{-\beta H(\mathbf{x})}}$$

这个结果可以用于计算玻尔兹曼分布的期望值, 例如, 可以计算 $x_i x_j x_k$ 的期望值:

$$\langle x_i x_j x_k \rangle = \frac{1}{Z} \sum_{\mathbf{x}} x_i x_j x_k e^{-\beta H(\mathbf{x})} = \frac{\sum_{\mathbf{x}} x_i x_j x_k e^{-\beta H(\mathbf{x})}}{\sum_{\mathbf{x}} e^{-\beta H(\mathbf{x})}}$$

将能量函数 $H(\mathbf{x})$ 代入上式可得：

$$\frac{\partial \log Z}{\partial J_{i,j,k}} = \langle x_i x_j x_k \rangle$$

因此，我们可以使用 $\frac{\partial \log Z}{\partial J_{i,j,k}}$ 来计算玻尔兹曼分布的期望值。

给定能量函数 $E(\mathbf{x}) = \frac{t\mathbf{x}^2}{2} + \frac{\mathbf{x}^4}{4}$ 和一个高斯密度函数 $Q_{\sigma}(\mathbf{x}) = \frac{\exp(-\frac{\mathbf{x}^2}{2\sigma^2})}{\sqrt{2\pi}\sigma}$ ，设定逆温度参数 $\beta = 1$ 。求 $Q_{\sigma}(\mathbf{x})$ 的 Gibbs 变分自由能 $G[Q_{\sigma}] = g(t, \sigma)$ 。

Gibbs 变分自由能是一个关于概率分布 $q(\mathbf{x})$ 的函数，定义为：

$$G[q] = \int q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x}$$

其中， $p(\mathbf{x})$ 是真实分布的概率密度函数， $q(\mathbf{x})$ 是近似分布的概率密度函数。

在这个问题中，真实分布的概率密度函数可以写成：

$$p(\mathbf{x}) = \frac{1}{Z} e^{-\beta E(\mathbf{x})} = \frac{1}{Z} e^{-\frac{t\mathbf{x}^2}{2} - \frac{\mathbf{x}^4}{4}}$$

其中， Z 是配分函数，定义为：

$$Z = \int e^{-\beta E(\mathbf{x})} d\mathbf{x} = \int e^{-\frac{t\mathbf{x}^2}{2} - \frac{\mathbf{x}^4}{4}} d\mathbf{x}$$

我们可以使用高斯密度函数 $Q_{\sigma}(\mathbf{x})$ 作为近似分布 $q(\mathbf{x})$ ，因此：

$$G[Q_{\sigma}] = \int Q_{\sigma}(\mathbf{x}) \log \frac{Q_{\sigma}(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x}$$

将 $p(\mathbf{x})$ 和 $Q_{\sigma}(\mathbf{x})$ 代入上式可得：

$$G[Q_{\sigma}] = \int Q_{\sigma}(\mathbf{x}) \log \frac{Q_{\sigma}(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x} = \int Q_{\sigma}(\mathbf{x}) (\log Q_{\sigma}(\mathbf{x}) - \log p(\mathbf{x})) d\mathbf{x} = \int Q_{\sigma}(\mathbf{x}) \left(\log Q_{\sigma}(\mathbf{x}) + \frac{t\mathbf{x}^2}{2} + \frac{\mathbf{x}^4}{4} - \log Z \right) d\mathbf{x} = \int Q_{\sigma}(\mathbf{x}) \left(\log Q_{\sigma}(\mathbf{x}) + \frac{t\mathbf{x}^2}{2} + \frac{\mathbf{x}^4}{4} \right) d\mathbf{x} - \log Z$$

其中，第一项是常数，可以忽略，第二项是近似分布的熵，可以用高斯分布的熵公式来计算：

$$-\int Q_{\sigma}(\mathbf{x}) \log Q_{\sigma}(\mathbf{x}) d\mathbf{x} = \frac{1}{2} \log(2\pi\sigma^2) + \frac{1}{2}$$

将其代入上式可得：

$$G[Q_{\sigma}] = \frac{t}{2\sigma^2} \int Q_{\sigma}(\mathbf{x}) \mathbf{x}^2 d\mathbf{x} + \frac{1}{4\sigma^4} \int Q_{\sigma}(\mathbf{x}) \mathbf{x}^4 d\mathbf{x} - \frac{1}{2}$$

我们可以使用高斯分布的均值和方差来计算上述两个积分。高斯密度函数 $Q_{\sigma}(\mathbf{x})$ 的均值为0，方差为 σ^2 ，因此：

$$\int Q_{\sigma}(\mathbf{x}) \mathbf{x}^2 d\mathbf{x} = \sigma^2 \int Q_{\sigma}(\mathbf{x}) d\mathbf{x} = \sigma^2 = 3\sigma^4$$

将其代入 $G[Q_{\sigma}]$ 的表达式可得：

$$G[Q_{\sigma}] = \frac{t}{2\sigma^2} \sigma^2 + \frac{1}{4\sigma^4} 3\sigma^4 - \frac{1}{2} = \frac{3t}{2} + \frac{3}{4\sigma^2} - \frac{1}{2}$$

因此， $Q_{\sigma}(\mathbf{x})$ 的Gibbs变分自由能为 $G[Q_{\sigma}] = \frac{3t}{2} + \frac{3}{4\sigma^2} - \frac{1}{2}$ 。这个结果表明，Gibbs变分自由能是 t 和 σ 的函数，可以用于比较不同的高斯分布 $q(\mathbf{x})$ 的近似能量。