10 Oct

变分自编码器 = 最小化先验分布 + 最大化互信息

By 苏剑林 | 2018-10-10 | 124988位读者 引用

这篇文章很简短，主要描述的是一个很有用、也不复杂、但是我居然这么久才发现的事实～

在《深度学习的互信息：无监督提取特征》一文中，我们通过先验分布和最大化互信息两个loss的加权组合来得到Deep INFOMAX模型最后的loss。在那篇文章中，虽然把故事讲完了，但是某种意义上来说，那只是个拼凑的loss。而本文则要证明那个loss可以由变分自编码器自然地导出来。

过程

不厌其烦地重复一下，变分自编码器（VAE）需要优化的loss是

$$\begin{aligned}&KL(\tilde{p}(x)p(z|x)\Vert q(z)q(x|z))\\ =&\iint \tilde{p}(x)p(z|x)\log \frac{\tilde{p}(x)p(z|x)}{q(x|z)q(z)}dzdx\end{aligned}\tag{1}$$

相关的论述在本博客已经出现多次了。VAE中既包含编码器，又包含解码器，如果我们只需要编码特征，那么再训练一个解码器就显得很累赘了。所以重点是怎么将解码器去掉。

其实再简单不过了，把VAE的loss分开两部分

$$\begin{aligned}&KL(\tilde{p}(x)p(z|x)\Vert q(z)q(x|z))\\ =&\iint \tilde{p}(x)p(z|x)\log \frac{p(z|x)}{q(z)}dzdx-\iint \tilde{p}(x)p(z|x)\log \frac{q(x|z)}{\tilde{p}(x)}dzdx\end{aligned}\tag{2}$$

第一项是先验分布的KL散度，第二项的$\log \frac{q(x|z)}{\tilde{p}(x)}$其实不也就是$x, z$的点互信息吗？假如$q(x|z)$具有无限的拟合能力，最终必然也会有$\tilde{p}(x)p(z|x)=q(x|z)p(z)$（贝叶斯公式），所以第二项也就是

$$KL(q(x|z)p(z)\|\tilde{p}(x)p(z)) = KL(\tilde{p}(x)p(z|x)\|\tilde{p}(x)p(z)) \tag{3}$$

就是x, z两个随机变量的互信息了，前面的负号意味着我们要最大化互信息。

剩下的处理过程就跟《深度学习的互信息：无监督提取特征》一样了，略。

结语

开头已经说了，这篇文章会很简短，没有什么内容。主要目的就是给出变分自编码器的loss的新理解（最小化先验分布 + 最大化互信息），然后就可以自然而言地导出Deep INFOMAX的loss。

如果我还没有写《深度学习的互信息：无监督提取特征》，那么我肯定会用这个出发点来讲解Deep INFOMAX，不过既然那篇文章都写了好几天了，所以只好另开这个简短的小文，来补充说明一下～

如果您需要引用本文，请参考：

苏剑林. (Oct. 10, 2018). 《变分自编码器 = 最小化先验分布 + 最大化互信息 》[Blog post]. Retrieved from https://spaces.ac.cn/archives/6088

```
@online{kexuefm-6088,
    title={变分自编码器 = 最小化先验分布 + 最大化互信息},
    author={苏剑林},
    year={2018},
    month={Oct},
    url={\url{https://spaces.ac.cn/archives/6088}},
}
```