

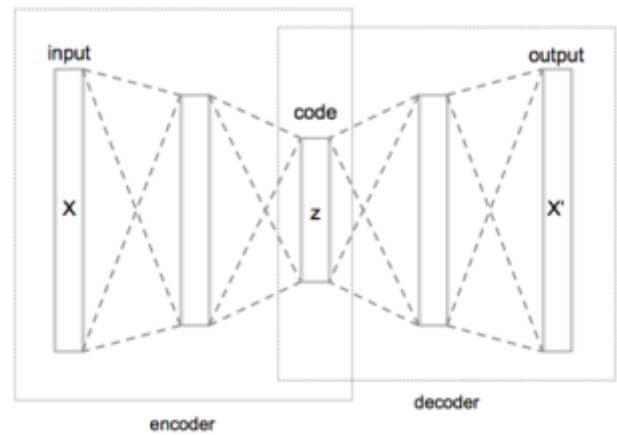
自编码器

维基百科，自由的百科全书

自编码器（英语：Autoencoder），也称自动编码器，是一种人工神经网络，在无监督学习中用于有效编码。自编码的目的是对一组数据学习出一种表示（也称表征，编码），通常用于降维。最近，自编码的概念广泛地用于数据的生成模型。^{[1][2]}自2010年以来，一些先进的人工智能在深度学习网络中采用了采用堆叠式稀疏自编码。^[3]

基本结构

自编码器有两个主要部分组成：编码器用于将输入编码，而解码器使用编码重构输入。



一个具有3个全连接隐藏层的自动编码器网络结构。

实现这个功能最简单的方式就是重复原始信号。然而，自编码器通常被迫近似地重构输入信号，重构结果仅仅包括原信号中最相关的部分。

自编码器的思想已经流行了几十年，其首次应用可以追溯到20世纪80年代。^{[4][5][6]}自编码器最传统的应用是降维或特征学习，现在这个概念已经推广到用于学习数据的生成模型。^{[1][2]}21世纪10年代的一些最强大的人工智能在深度神经网络中采用了自编码器。^[3]

最简单的自编码器形式是一个前馈的、非循环的神经网络，用一层或多层隐藏层链接输入和输出。输出层节点数和输入层一致。其目的是重构输入（最小化输入和输出之间的差异），而不是在给定输入的情况下预测目标值，所以自编码器属于无监督学习。

最简单的自编码器形式是一个前馈的、非循环的神经网络，类似于多层感知器（MLP）中的单层感知器，用一层或多层隐藏层链接输入和输出。输出层具有与输入层相同数量的节点（神经元）。输出层节点数和输入层一致。其目的是重构输入（最小化输入和输出之间的差异），而不是在给定输入 $\mathbf{X}$ 的情况下预测目标值 $\mathbf{Y}$ ，所以自编码器属于无监督学习。

自编码器由编码器和解码器组成，二者可以被分别定义为变换 ϕ 和 ψ ，使得：

$$\begin{aligned} \phi &: \mathcal{X} \rightarrow \mathcal{F} \\ \psi &: \mathcal{F} \rightarrow \mathcal{X} \\ \phi, \psi &= \arg \min_{\phi, \psi} \|\mathcal{X} - (\psi \circ \phi) \mathcal{X}\|^2 \end{aligned}$$

在最简单的情况下，给定一个隐藏层，自编码器的编码阶段接受输入 $\mathbf{x} \in \mathbb{R}^d = \mathcal{X}$ 并将其映射到 $\mathbf{h} \in \mathbb{R}^p = \mathcal{F}$ ：

$$\mathbf{h} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b})$$

像 $\mathbf{h}$ 通常表示编码、潜变量或潜在表示。 σ 是一个逐元素的激活函数（例如sigmoid函数或线性整流函数）。 $\mathbf{W}$ 是权重矩阵， $\mathbf{b}$ 是偏置向量。权重和偏置通常随机初始化，并在训练期间通过反向传播迭代更新。自编码器的解码阶段映射 $\mathbf{h}$ 到重构 $\mathbf{x}'$ （与 $\mathbf{x}$ 形状一致）：

$$\mathbf{x}' = \sigma'(\mathbf{W}'\mathbf{h} + \mathbf{b}')$$

其中解码器部分的 σ' , $\mathbf{W}'$, $\mathbf{b}'$ 可能与编码器部分的 σ , $\mathbf{W}$, $\mathbf{b}$ 无关。

自编码器被训练来最小化重建误差（如平方误差），通常被称为 "损失"：

$$\mathcal{L}(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|^2 = \|\mathbf{x} - \sigma'(\mathbf{W}'(\sigma(\mathbf{W}\mathbf{x} + \mathbf{b})) + \mathbf{b}')\|^2$$

其中 $\mathbf{x}$ 通常在训练集上平均。

如前所述，和其它前馈神经网络一样，自编码器的训练是通过误差的反向传播进行的。

当特征空间 $\mathcal{F}$ 的维度比输入空间 $\mathcal{X}$ 低时，特征向量 $\phi(\mathbf{x})$ 可以看作时输入 $\mathbf{x}$ 的压缩表示，这就是不完备自动编码（undercomplete autoencoders）的情况。如果隐藏层大于（过完备）或等于输入层的数量，或者隐藏单元的容量足够大，自编码器就可能学会恒等函数而变得无用。然而，实验结果表明过完备自编码器（overcomplete autoencoders）仍然可能学习到有用的特征。^[7]在理想情况下，编码的维度和模型容量可以根据待建模数据分布的复杂性来设定，采用这种方式的一种途径是正则化自编码器。^[4]

另見

- [表征学习](#)

參考

1. Auto-Encoding Variational Bayes, Kingma, D.P. and Welling, M., ArXiv e-prints, 2013 arxiv.org/abs/1312.6114 (<https://arxiv.org/abs/1312.6114>)
2. Generating Faces with Torch, Boesen A., Larsen L. and Sonderby S.K., 2015 torch.ch/blog/2015/11/13/gan.html (<http://torch.ch/blog/2015/11/13/gan.html>)
3. Domingos, Pedro. 4. The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Basic Books. 2015. "Deeper into the Brain" subsection. ISBN 978-046506192-1.
4. Bengio, Yoshua; Courville, Aaron. Deep learning. *Deep Learning*. Cambridge, Massachusetts. 2016 [2022-06-06]. ISBN 0-262-03561-8. OCLC 955778308. （原始内容存档于2021-02-17） .
5. Schmidhuber, Jürgen. Deep learning in neural networks: An overview. *Neural Networks*. January 2015, **61**: 85–117. PMID 25462637. S2CID 11715509. arXiv:1404.7828 . doi:10.1016/j.neunet.2014.09.003.
6. Hinton, G. E., & Zemel, R. S. (1994). Autoencoders, minimum description length and Helmholtz free energy. In *Advances in neural information processing systems 6* (pp. 3-10).
7. Bengio, Y. Learning Deep Architectures for AI (PDF). *Foundations and Trends in Machine Learning*. 2009, **2** (8): 1795–7 [2022-06-06]. CiteSeerX 10.1.1.701.9550 . PMID 23946944. doi:10.1561/22000000006. （原始内容 (PDF)存档于2015-12-23） .

本页面最后修订于2022年7月16日 (星期六) 03:10。

本站的全部文字在知识共享 署名-相同方式共享 3.0协议之条款下提供，附加条款亦可能应用。（请参阅使用条款）
Wikipedia®和维基百科标志是维基媒体基金会的注册商标；维基™是维基媒体基金会的商标。
维基媒体基金会是按美国国内稅收法501(c)(3)登记的非营利慈善机构。