

# Machine Learning Engineer Nanodegree

## Capstone Proposal

JinXian Dong

March 20th, 2018

### 项目背景

猫狗识别，源于kaggle竞赛，属于计算机视觉的研究领域。猫狗识别项目，涉及到图片分类及目标识别到内容。2012年开始使用扩展了深度的CNN在ImageNet竞赛中获得极佳表现，分类能力和识别能力均大幅度增加，并取得突破性进展<sup>[1]</sup>。猫狗识别项目中，涉及到的分类问题和目标识别问题基于CNN完全可以得到解决。

我个人选择这个项目，也是基于上述，图像分别、目标识别是机器学习中的经典问题。出于学习的考虑，自然想一探究竟。

### 问题描述

我将要解决的问题是，对于未知的图片输入，判断图片中是否有猫狗，并区分出具体是猫还是狗。这个问题，我所知道的，至少有两种解决办法：一是传统的监督学习，如支持向量机等，二是常用的深度学习，如卷积神经网络。我将通过已有的数据集，训练模型，并通过验证集测试识别结果。这个地方给出具体的识别概率，因此这个问题可衡量，可量化。然后通过现实生活中的未知猫狗图片，进行输入识别，模型对于相同的输入和会输出特定的识别率，可重现。

### 输入数据

数据集来源于Kaggle，原数据集有12500只猫和12500只狗。图片的命令都是type.num.jpg，图片的大小有不同的尺寸，图片都是彩色的，包含RGB三个通道数据。

另外，除了25000张已作出标记的训练数据，还有12500张未作标记的数据集。

### 解决办法

首先传统的机器学习方法，对于此类问题，也有一定的处理能力。但是业界公认的解决方案，是卷积神经网络CNN（2012后，对于此类问题解决能力的提升，均是采用了更加优化设计的CNN模型）。因此我们采用CNN，作为本项目的解决方案。

CNN的基本结构由输入层、卷积层、取样层、全连接层及输出层构成。卷积层和取样层一般会取若干个，采用卷积层和取样层交替设置，即一个卷积层连接一个取样层，取样层后再连接一个卷积层，依此类推。由于卷积层中输出特征面的每个神经元与其输入进行局部连接，并通过对应的连接权值与局部输入进行加权求和再加上偏置值，得到该神经元输入值，该过程等同于卷积过程，卷积神经网络也由此而得名<sup>[2]</sup>。

### 基准模型

我们将竞赛中比较经典的一些CNN模型作为基准模型，如：AlexNet模型、GoogLeNet模型与VGG模型。

- AlexNet模型，由Krizhevsky等人首次应用，主要是通过加深CNN模型的深度，及采用了Relu激活函数和dropout技术。该模型包含5个卷积层和2个全连接层。
- GoogLeNet模型，由Szegedy等人提出，通过极大的增加CNN的深度，提高准确率。

- VGG模型，则是用具有小卷积核的多个卷积层替换一个具有较大卷积核的卷积层。

## 评估指标

直接采用kaggle官方提供的公式，其最终得分越低，效果越好，如下：

$$\text{LogLoss} = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)],$$

- 说明:
  - $n$ 是测试集的图片数量
  - $\hat{y}_i$ 是猜测图片是狗的概率
  - $y_i$ 为1的时候图片为狗，0的时候图片为猫
  - $\log()$ 是自然对数

## 设计大纲

本项目的遵循传统CNN模型训练和构建的一般步骤：

1. 输入层处理，主要是调整图片的大小即归一化处理，有可能也需要灰度处理，向量化等。
2. CNN模型结构的设计，这里会参考以前比赛中的经典基准模型。一般是卷积层的堆叠和深度的考量。由于我的计算资源有限，这个过程中，我可能会使用一些已经训练好的模型，作迁移学习。
3. 最后我会观察，测试，我选取的设计方案，在验证数据上的表现，调整超参，以达标要求。完成项目。

## 参考文献

- [1] Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton. Imagenet classification with deep convolutional neural networks//Proceedings of Advances in Neural Information Processing Systems, Lake Tahoe,USA , 2012:1097-1105
- [2] Yann LeCun, Leon Bottou, Yoshua Bengio, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11):2278-2324