

**AI**

**课**

**程**

**报**

**告**

**姓名 ：姚菁**

**学号 ：2015202031**

**完成时间 2017年12月13日**

**目 录**

**【课程内容整理】………………………………………………………3**

**A、Fundamentals of Machine Learning(Traditional Inelligence)……………3**

**B、Neural Networks and Deep Learnig（Deep Intelligence）………………4**

**【课程实验总结】………………………………………………………5**

**A、第一阶段……………………………………………………………………5**

**B、第二阶段……………………………………………………………………5**

**C、第三阶段……………………………………………………………………6**

**【课外资料查阅】**………………………………………………………**6**

**A、BP（Backforword Propagation）算法……………………………………6**

**B、Training Tricks……………………………………………………………7**

**【课程内容整理】**

1. **Fundamentals of Machine Learning(Traditional Inelligence)**

**Chapter 1:** The Machine Learning Landsape---- i）Machine-learning的定义，数据驱动是它与传统编程方法的最大区别。 Ii）Machine-learning的分类，可根据多方面的标准进行划分，eg：supervision、semisupervised、reinforcement learning； online、 batch learning等。 Iii）Machine-learning面临的挑战，主要包括两个方面：一是数据的数量不足质量不好，二是算法容易导致overfitting和underfittng的问题。

**Chapter 2：**End-to-End Machine Learning Project----介绍了通过machine learning的方法来解决一个实际问题的完整流程以及训练的小trick。

1. Look at the big picture.

2. Get the data.

3. Discover and visualize the data to gain insights.

4. Prepare the data for Machine Learning algorithms.

5. Select a model and train it.

6. Fine-tune your model.

7. Present your solution.

8. Launch, monitor, and maintain your system.

Tricks：i）通过RMSE或者MAE来衡量算法优劣 ii）可视化观察数据，预处理 iii）网格搜索、随机搜索来调整参数，通过交叉验证进行选择。

**Chapter 3：**Classification----运用经典数据集Mnist为基础来进行说明，包括Binary-classifier、multiclass-claasifier、multilabel-classification、multioutput-classification。

**Chapter 4：**Trainig models----i）两种计算参数的方法 Normal Equition和Gradient Descend各自的优劣以及适应情形，其中GD则还分为Batch GD，Mini-batch GD, and Stochastic GD，各有优劣。 Ii）多项式回归，在线性回归的基础上增加次方向以便更好地拟合数据，但可能会造成overfitting，可以通过regularize来解决包括Ridge Regression、Lasso Regression 和 Elastic Net。 Iii）两种用于分类的回归方法 Logistic Regression和Softmax Regression。

**Chapter 5：**Support Vector Machines----可用于线性分类、非线性分类、线性回归等，尤其适用于中小型数据集的复杂分类问题。其中具体包括Large margin Classification、soft Margin classification、nonlinear svmclassificaitn、 svm regression、 quadratic programming等

**Chapter 6：**Decision Trees----通过构建决策树来解决分类和回归问题，可以使用Gini Impuroty、 entropy或者MSE来衡量模型的优劣。

**Chapter 7：**Ensemble learning and Random Forest----i）Ensemble learning是通过合并多个predictor的预测结果或者综合考虑多个predictor的预测结果从而得出最终结果的方法，具体有voting classifiers、Bagging and Pasting、Boosting、Stacking等，Random Forest也是其中的一种算法。

**Chapter 8：**Dimensionality reduction----由于高维向量数据训练难度大，因此需要降维以加快训练速度、优化训练结果。I）主要方法包括Projection和Manifold learning。 Ii）经典的算法有PCA（Principal Component Analysis）、IPCA、KPCA以及LLE（locally linear embedding）

**B．Neural Networks and Deep Learnig（Deep Intelligence）**

**Chapter 9：**Up and Running with Tensorflow----主要介绍了tensorflow的一些入门基础知识，从环境安装到具体实现。I）Tensorflow编程包括两个部分，一是创建一个Computational Graph，二是在Session中运行。 Ii）如何利用Tensorflow来实现一些基础的机器学习算法，包括回归、梯度下降、最优化等。 Iii）Tensorboard的使用。

**Chapter 10：**Introduction to Artificial Neural Networks----i）ANN发展成熟的原因归结为可训练数据的大量增加、计算机性能的提高和训练算法自身的优化。 Ii）从最简单的Perceptron开始，引入hidden layer和DNN，并介绍了DNN的训练方法BP。

**Chapter 11：**Trainig Deep Neural Nets----i）Vanishing/Exploding Gradients Problems的解决方法：参数初始化Xavier initialization，Relu->LeakyRelu/ELU，Batch normalizaiotn，Grdient Clipping等。Ii）Faster Optimizers，快于普通的梯度下降，包括Momentum optimization、 Nesterov Accelerated Gradient， AdaGrad， RMSProp， Adam optimizer。 Iii）为了避免过拟合所采用的一些Regularization的方法，Early stopping、ℓ1 and ℓ2 Regularization、Dropout、Max-Norm Regularization、Data Augmentation。

**Chapter 13：**Convolutional Neural Networks----与人脑视觉皮层的工作原理相关，主要用于图像识别领域。I）The Architecture of the Visual Cortex，利用视觉皮层的神经元局有局部感受野的原理。 Ii）Convolutional Layer，每一层的每个神经元与前一层的一个小长方形区域连接，其中的参数矩阵为filters，大小为[*fh, fw, fn, fn′]。*iii）Pooling Layer，区域也为一个小长方形区域，目的是为了减轻内存和计算的负担，常用max-pooling。iiii）经典的CNN模型包括LeNet-5、AlexNet、GoogleNet、ResNet。

**Chapter 14:** Recurrent neural networks----RNN是一种能够利用已有的信息预测将来的网络，能够分析时间序列，包含循环并且允许信息的持久化。本章主要介绍了RNN的一些基本概念以及面临的问题。

**【课程实验总结】**

1. **第一阶段**

**参与任务：**负责小车组件的购买，与小组成员一起实现小车的组装和接线，阅读、理解并且将程序烧如Arduino板，与畅雅雯同学一起制作展示PPT。

**心得体会：**通过本次课程试验将纯粹的机器编程与实际动手操作很好地结合起来了，在实验中不仅需要理解控制小车运动的程序逻辑，还需要完成小车的实际接线等工作。这首先需要我们理解各个工作组件之间的联系以及各个引脚的具体作用，同时与代码的控制逻辑相结合才能够完成初步的接线工作；第一次完成接线工作，希望小车能够实现初步的行驶，但是这中间却出现了各种各样的问题，我们通过反复地调整接线方式，最后联通充电宝才能够让小车初步地动起来。但是由于零件自身性能的原因，两侧马达的转速非常不均，导致小车不能直线行驶，我们考虑通过后期程序调整两轮转速来解决；第二步我们将程序烧录到Arduino板中。

1. **第二阶段**

**参与任务：**重新接线实现避障，实现手机-电脑图片抓取，制作展示PPT，课堂展示。

**心得体会：**此次试验我们吸取了第一次实验的教训，并且打算在第一次试验的基础上进行改进。由于第一次试验中出现了许多不明原因的问题，因此我们决定重新进行一次连线，并且测试马达性能寻找两个转速最接近的马达，虽然这些工作对功能上没有很大的帮助，但还是在很大程度上促进了我们进一步的实验，因此在实验中我们一定不能忽略这些看似不重要的细节对我们的影响。最后，我通过学习DroidCam的使用方法，在同一局域网或蓝牙连接的情况下实现了手机拍摄与电脑之间的实时传送，为下一阶段更高程度的功能打下基础。

1. **第三阶段**

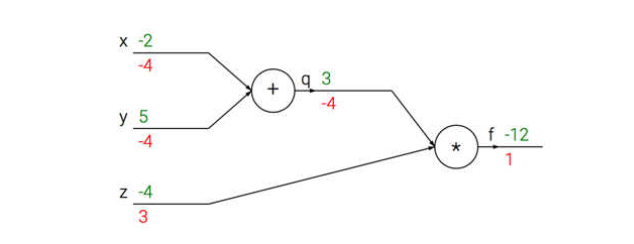
**参与任务：**手机电脑照片抓取、图像识别功能实现、制作展示PPT。

**心得体会：**前两个阶段我们组的工作都主要集中在对小车自身性能的改进上，在第三阶段又进一步加入了一些人工智能，通过机器学习的算法来进行实现。而通过书写这个功能代码也让我对于机器学习、图像识别和分类器的使用有了更具体的认识，我们主要利用的是tensorflow已经实现的对mnist数据集的图片进行分类的模型——softmax分类，用交叉熵损失来衡量模型的优劣，通过GD来进行最优化，最终得到准确度约为95%的模型，加上我们自己对于照片的处理和输入模型的接口来实现识别手写数字图片的功能。其中主要需要进行的工作则是数据的预处理，由于mnist数据集中的数据拥有自己特殊的格式，因此我们的输入数据也要进行相同的预处理才能达到较高的准确率，可见一个好的数据对于算法的重要性。其次，也加深了我对分类算法softmax，交叉熵损失，梯度下降，最优化和图像识别常用网络卷积神经网络CNN等的理解和认识。

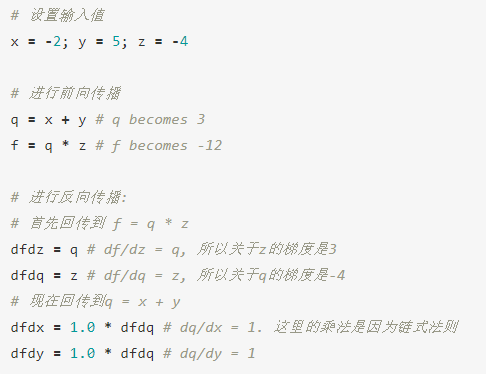
**【课外资料查询】**

1. **BP（Backforword Propagation）算法**

由于对BP算法没有理解透彻，因此查阅相关资料如下。BP算法是利用链式法则递归计算表达式的梯度的方法，理解BP算法需要有一定的微积分知识。直观地将每个计算都看做一个门单元，那么在整个计算线路中，每个们单元都会得到一些输入并且立即计算：1.这个门的输出值；2.门的输出值关于输入值的局部梯度。第一部分用于前向传播计算输出值，第二部分则用于在回传过程中计算门单元对于输入值的梯度。例：



f(x,y,z) = (x+y)\*z: q=x+y, f = qz;



反向传播是从尾部的输出只开始向前计算的，可以认为梯度是从计算链路中回流。

1. **Training Tricks**

虽然现在神经网络的发展已经比较成熟了，但是仍然问题和瓶颈大于成果，尤其是在很多网络的训练过程中非常容易出现underfitting和overfitting的问题，因此我查阅了相关资料，找到了一些有经验的前辈关于训练神经网络的小Trick的分享。

1、参数初始化。一定要对参数做初始化，否则可能会减慢收敛速度，影响手链结果，甚至造成Nan等一系列的问题。有Xavier初始化法和He初始化法，并且可以采用uniform均匀分布初始化和normal高斯分布初始化。

2、数据预处理方式。包括常用的zero-center ,X -= np.mean(X, axis = 0) # zero-center。  
标准化X /= np.std(X, axis = 0) # normalize。

3、梯度归一化,即算出来的梯度除以minibatch size。

Gradient-clip(梯度裁剪)：限制最大梯度,其实是value = sqrt(w1^2+w2^2….),如果value超过了阈值,就算一个衰减系系数,让value的值等于阈值: 5,10,15。

Dropout：对小数据防止过拟合有很好的效果,值一般设为0.5,小数据上dropout+sgd，位

adam,adadelta等,在小数据上,实验效果不如sgd, sgd收敛速度会慢一些，但是最终收敛后的结果一般都比较好。如果使用sgd的话,可以选择从1.0或者0.1的学习率开始,隔一段时间,在验证集上检查一下,如果cost没有下降,就对学习率减半，也可以先用ada系列先跑,最后快收敛的时候,更换成sgd继续训练.同样也会有提升.据说adadelta一般在分类问题上效果比较好，adam在生成问题上效果比较好。

除了gate之类的地方,需要把输出限制成0-1之外,尽量不要用sigmoid,可以用tanh或者relu之类的激活函数.1. sigmoid函数在-4到4的区间里，才有较大的梯度。之外的区间，梯度接近0，很容易造成梯度消失问题。2. 输入0均值，sigmoid函数的输出不是0均值的。

rnn的dim和embdding size,一般从128上下开始调整. batch size,一般从128左右开始调整.batch size合适最重要,并不是越大越好.

尽量对数据做shuffle

4、Batch Normalization

5、Ensemble：深度学习中一般有以下几种方式

同样的参数,不同的初始化方式

不同的参数,通过cross-validation,选取最好的几组

同样的参数,模型训练的不同阶段，即不同迭代次数的模型。

不同的模型,进行线性融合. 例如RNN和传统模型。

希望在训练网络的过程中也能够自己总结经验，不断提升网络训练的结果。