机器学习经验总结和资源

# 数据预处理

## 数据转换

### log转换

### 标准化转换

### 衍生特征

# 深度学习

## 初始参数设置

### 参考资源（出处）

Deep Learning with Tensorflow (2nd)[ <https://www.packtpub.com/big-data-and-business-intelligence/deep-learning-tensorflow-second-edition>]

### 网络架构参数

Deep Learning with Tensorflow (2nd)[ <https://www.packtpub.com/big-data-and-business-intelligence/deep-learning-tensorflow-second-edition>]

**setting the hidden layer configuration using just two rules:**

* The number of hidden layers equals one
* The number of neurons in that layer is the mean of the neurons in the input and output layers

#### --层数：N（l）=N(s)/{a\*[N（i）+N（o）]

---N（i）：输入层神经元数量

---N（o）：输出层神经元数量（如果有n个类别，则为n）

---N（s）：训练样本数

#### --每层的神经元数:N(n)= [N（i）+N（o）]/2

### 初始化

#### Weight初始化：不能全部初始化为0

#### Bias初始化：首先考虑全为0，或者比较小的值0.01

### 优化器

#### AdamOptimizer

most of the cases, we can utilize Adam, it works well and without having to think much about learning rates and so on.

#### RMSPropOptimizer

The suggested setting value of the decay parameter is 0.9, while a good default value for the learning rate is 0.001:

optimizer = tf.train.RMSPropOptimizer( 0.001, 0.9). minimize( cost\_op)

#### MomentumOptimizer

Researchers also recommend using Momentum optimizer while training a deep network such as CNN.

#### 同时参考

一文看懂各种神经网络优化算法：从梯度下降到Adam方法

<https://zhuanlan.zhihu.com/p/27449596?utm_source=weibo&utm_medium=social>

## 超参搜索/优化

### GridSearchCV

(http://scikit-learn.org/stable/modules/generated/sklearn.model\_selection.GridSearchCV.html# sklearn.model\_selection.GridSearchCV) exhaustively considers all parameter combinations

### RandomizedSearchCV

(http:// scikit-learn.org/ stable/ modules/ generated/ sklearn.model\_selection.RandomizedSearchCV.html# sklearn.model\_selection.RandomizedSearchCV) can sample a given number of candidates from a parameter space with a specified distribution.

### Keras deep learning

#### How to Grid Search Hyperparameters for Deep Learning Models in Python With Keras

**<代码实现：keras\_sklearn\_GridSearchHyperparameters.py >**

<https://machinelearningmastery.com/grid-search-hyperparameters-deep-learning-models-python-keras/>

1. How to use Keras models in scikit-learn.
2. How to use grid search in scikit-learn.
3. How to tune batch size and training epochs.
4. How to tune optimization algorithms.
5. How to tune learning rate and momentum.
6. How to tune network weight initialization.
7. How to tune activation functions.
8. How to tune dropout regularization.
9. How to tune the number of neurons in the hidden layer.