**Deep learning material**

Neural network questions:

1. Loss function的选择:

Logistic loss (我们需要算probability不只是0或1)

cross entropy for soft max（soft max是占总体百分比，我们各个类别总和不为1）

1. activation function的选择:

sigmoid (vanishing gradient problem是否不同layer需要用不同activation fuction)

ReLU (parameter learned by gradient descent是否必须和gradient descent合用)

1. 基本参数的选择：

number hidden layers（flatter is better）

number of neurons (Ng说通常每层neuron数一致)

batch size （如何确定batch size？）

number of epochs (early stopping 感觉可以从小试起)

optimization method (gradient descent, LG, GS, contract descent, etc.)

Adaptive learning rate (Adagrad 如果不用gradient descent效果如何, momentum如何设定常数值)

Avoid overfitting:

1. early stopping
2. weight decay（是否可以和dropout合用？）
3. dropout（被dropout部分neuron testing时theta如何决定？）
4. cross validation method (k-fold 具体split比例如何决定？)

Deep learning network:

1. CNN (subsampling: 动物，汽车等显著特征)
2. RNN (p222: video caption; p289: acoustic features 是否通过识别语音生成字幕快速甄别视频类别？)
3. Ultra deep network (FractalNet, residual network, highway network)

**Youtube-8M-paper**

Classes:

4800 entities from YouTube video annotation system, 24 top-level categories

Models:

a). frame-level

1. Frame-level model and average pooling

Train 4800 independent 1 vs all classifiers to all entities, and then aggregate the score for annotations to video-level score using simple average

1. Deep Bag of Frame (DBoF) pooling (regardless of sequence of frames)

Layers: input layer (N\*k) 🡪 fully connected layer (M, ReLU) 🡪 max pooling layer 🡪 … fully connected layer … 🡪 output layer (Logistic or Softmax)

Optimization: Stocastic Gradient Descent (SGD)

Loss function: logistic loss for logistic layer and cross-entropy loss for softmax layer

1. Long Short-Term Memory (LSTM) (include the effect of frame sequence)

2 layers with 1024 units, linearly increasing per-frame weights from 1/N to 1 for the last frame

b). video-level

1. Logistic Regression
2. Hinge Loss
3. Mixture of Experts (MoE)

Perform very well for multi-modal classification

Conclusion:

In terms of frame-level models, LSTM performs best;

In terms of video-level models, MoE performs best;

MoE and LSTM perform almost the same because the underlying frame-level features are already very strong