

---

# **A2 homework submission**

## **Team: the gurecki**

### **Deep Learning 2015, Spring**

---

**David Halpern**  
Department of Psychology  
New York University  
david.halpern@nyu.edu

**Anselm Rothe**  
Department of Psychology  
New York University  
ar3918@nyu.edu

**Alex Rich**  
Department of Psychology  
New York University  
asr443@nyu.edu

## **1 Input**

## **2 Architecture**

### **2.1 Stage 1**

### **2.2 Stage 2**

### **2.3 Stage 3**

## **3 Learning Techniques**

## **4 Learning Procedure**

## **5 Architecture**

### **5.1 Input**

The input is a 3D array with three 2D feature maps of size  $32 \times 32$ . We preprocessed the training data by normalizing across features so that each feature had a mean of 0 and a standard deviation of 1 and then used the same normalization for the test data. The data was also normed per pixel so each pixel across images had a mean of 0 and a standard deviation of 1. Each image is padded with zeros of size  $2 \times 2 \times 2 \times 2$  in order to make for better convolutional filter learning.

### **5.2 Model**

The first layer applies  $23 \times 7 \times 7$  convolutional filters with a stride of 2. These filters have a ReLU activation function and are fed into a max pooling layer which takes the max of each  $3 \times 3$  region with a step size of  $2 \times 2$ . Each filter is now  $22 \times 22$ . This they are all fed into a linear layer with 50 outputs nodes. The outputs are then run through a Tanh activation function and finally put through another linear layer with 10 output nodes, one for each category. The model then uses log softmax to convert energies into choices.

## 6 Learning Techniques

We used dropout to train the convolutional layers.

## 7 Training Procedure

We used the to train our model. The model was trained using stochastic gradient descent with a learning rate of .001 ???, a batch size of 128 a momentum of .9 and weight decay of ???. We used the typical classNLLcriterion as the loss function. We used a validation set of 500 of the original 5000 labeled training examples. The optimization procedure was run for 100 epochs over the remaining 4500 training data.

Performance ???

## 8 Experiments

We attempted to use uniform prescription to train our model using the unlabeled data. In order to do this, we used KL Distance as our loss function so that the target for the unlabeled data could be a uniform distribution on all of the categories. The targets for the labeled data were a delta function with all probability mass on the true category. We tried using various amounts of unlabeled data from 500 to all 100000 but this did not seem to improve performance

## References

- [1] Dosovitskiy, A., Springenberg, J. T., Riedmiller, M., & Brox, T. (2014). Discriminative unsupervised feature learning with convolutional neural networks. *Advances in Neural Information Processing Systems*, pp. 766-774.
- [2] Xiang Zhang (2015). *How to Train STL-10 Well*. <http://goo.gl/xJGvyH>