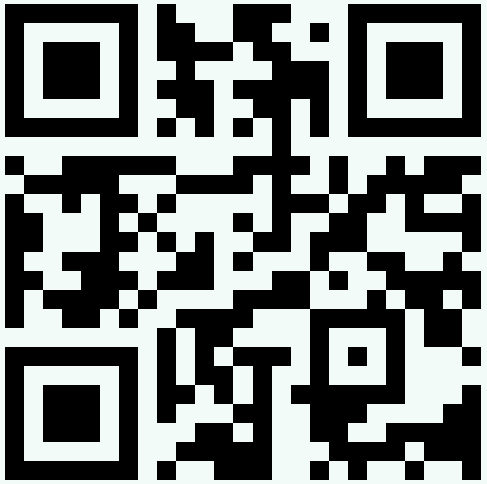
CSCI 567 discussion: Neural Nets

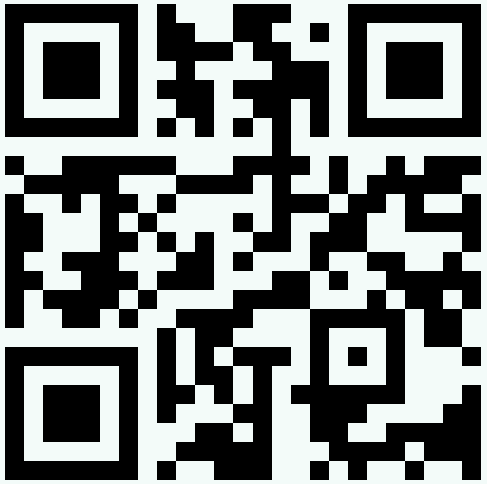


Yue Wu
<wu.yue@usc.edu>

Today

- Practice: 5 multiple choice questions
- Neural nets, CNN

- Demonstration of back propagation



1.1. Which of the following is/are **true** about neural nets? **BC**

(A) A neural net with one hidden layer and a fixed number of neurons can approximate any continuous function.

(B) A fully connected feedforward neural net without nonlinear activation functions is the same as a linear model.

(C) Dropping random neurons in each iteration of Backpropagation helps prevent overfitting.

(D) A max-pooling layer with a 2×2 filter has 4 parameters to be learned. → 0 params

(A) universal approx. thm. : #neuron → ∞. X

(B) Feedforward:

$$y = h_n(W_n h_{n-1}(\cdots h_2(W_2 h_1(W_1 x))\cdots))$$

If no nonlinear: → $W = W_n \cdots W_1$

$$y = W_n W_{n-1} \cdots W_1 x$$ one linear layer

1.2. Which of the following can help prevent overfitting in neural nets?

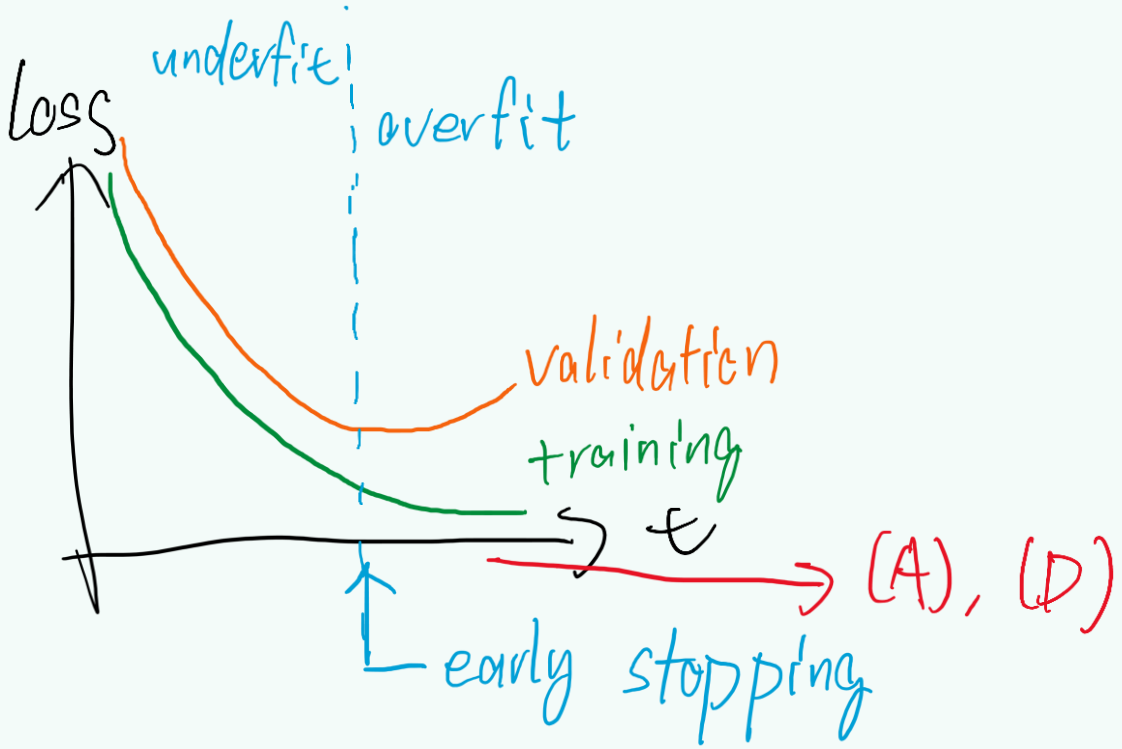
(A) Retraining on the same data many times.

(B) Using a validation set for early stopping.

(C) Data augmentation.

(D) Training until you get the smallest training error.

BC



1.3. Suppose a convolution layer takes an 8×8 image with 3 channels as input and outputs a $4 \times 4 \times 8$ volume. Which of the following is a possible configuration of this layer?

(A) One 4×4 filter with depth 8, stride 2, 1 pixel of zero-padding.
(B) Three 4×4 filters with depth 8, stride 2, no zero-padding.
(C) Eight 4×4 filters with depth 3, stride 2, 1 pixel of zero-padding.
(D) Eight 4×4 filters with depth 3, stride 1, 1 pixel of zero-padding.

C

8X8X3 ⟶ 4X4X8

depth=3 channels = 8

8X8 —pad=1→ 10X10 —stride=2, W=H=4→ 4X4

1.4. How many parameters do we need to learn for the following network structure? An $32 \times 32 \times 3$ image input, followed by a convolution layer with 3 filters of size 3×3 (stride 1, 1 pixel of zero-padding), then another convolution layer with 4 filters of size 2×2 (stride 2, no zero-padding), and finally a max-pooling layer with a 2×2 filter (stride 1, no zero-padding). (Note: the depth of all filters are not explicitly spelled out, and we assume no bias/intercept terms are used.)

(A) 43
(B) 97
(C) 129
(D) 145

C. $3 \times (3 \times 3 \times 3) + 4 \times (2 \times 2 \times 3)$

convolution layer: # parameter doesn't depend on image size

1.4. How many parameters do we need to learn for the following network structure? An $32 \times 32 \times 3$ image input, followed by a convolution layer with 3 filters of size 3×3 (stride 1, 1 pixel of zero-padding), then another convolution layer with 4 filters of size 2×2 (stride 2, no zero-padding), and finally a max-pooling layer with a 2×2 filter (stride 1, no zero-padding). (Note: the depth of all filters are not explicitly spelled out, and we assume no bias/intercept terms are used.)

1.5. What is the final output dimension of the last question?

(A) $15 \times 15 \times 1$
(B) $16 \times 16 \times 4$
(C) $32 \times 32 \times 1$
(D) $15 \times 15 \times 4$

D

common filter size:

3x3, padding 1
5x5, padding 2
7x7, padding 3
} preserve w, h

Convolutional neural networks (ConvNets/CNNs) | Architecture

Summary for convolution layer

Input: a volume of size $W_1 \times H_1 \times D_1$

Hyperparameters:

- K filters of size $F \times F$
- stride S
- amount of zero padding P (for one side)

Output: a volume of size $W_2 \times H_2 \times D_2$ where

- $W_2 = (W_1 + 2P - F)/S + 1$
- $H_2 = (H_1 + 2P - F)/S + 1$
- $D_2 = K$

#parameters: $(F \times F \times D_1 + 1) \times K$ weights

43 / 47

32x32

conv w=h=3, S=1, P=1

$\Rightarrow \frac{32+2-3}{1} + 1 = 32$

conv w=h=2, S=2, P=0

$\Rightarrow \frac{32+0-2}{2} + 1 = 16$

max pool w=h=2 S=1

$\Rightarrow \frac{16+0-2}{1} + 1 = 15.$

Back propagation demo

- https://xnought.github.io/backprop-explainer/