Friday, February 12, 2021 3:07 PM

Lecture 8

Regularization

Many methods have been developed to reduce overfitting during training

- General idea: overfitting results from many parameters working together to overfit data
- Some commonly used methods
  - weight regularization / penalty
    - loss += $\varepsilon \sum \|w\|^2$ "L2 weight regularization" (biases too)
    - $\varepsilon \sum \|w\|$ "L1 " "
  - Dropout - randomly drops activations (nodes) during training. → restored at evaluation time

Friday, February 12, 2021 4:04 PM

Convolutional Neural Networks

DNN flattens image → vector, doesn't preserve locality.

CNN preserve 2d structure → 

rather than learn giant matrix

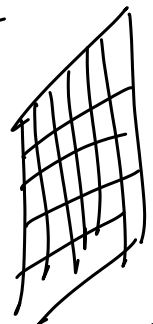
learn much smaller "filter" matrix

local feature map

idea: drag filter across image, each filter learns 1 "feature"

produce 2d maps of features in image

6x6



3x3 filter (w_{ab}, b) $a,b = 1,2,3$

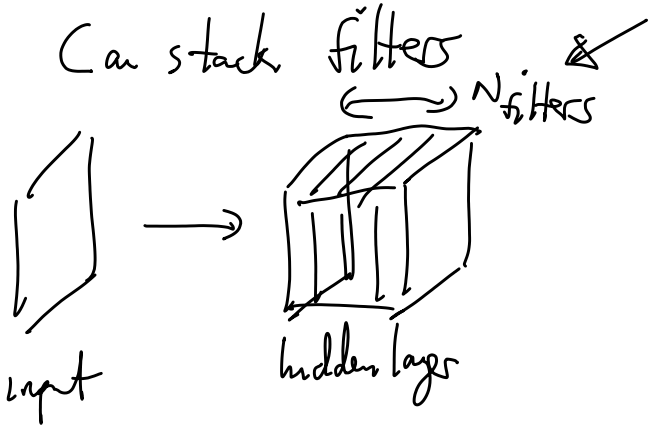
$$A\left(\sum_{a,b=1}^{3} x_{i+a,j+b}\, w_{ab} + b\right) = x^{(1)}_{i,j}$$

produces hidden layer which is an image!

x_{ij} $i,j = 1, \dots N_{pixels}$

Friday, February 12, 2021 4:21 PM

Can stack filters



"Convolutional layer"

Can stack conv. layers!

$X^{(l)}_{i,j,k}$ (i: pixels, j: pixels, k: channel)

$$A\left(\sum_{a,b,c} X^{(l)}_{i+a,j+b,k+c} W_{abc} + b\right)$$

Pooling layer — reduce image size drastically

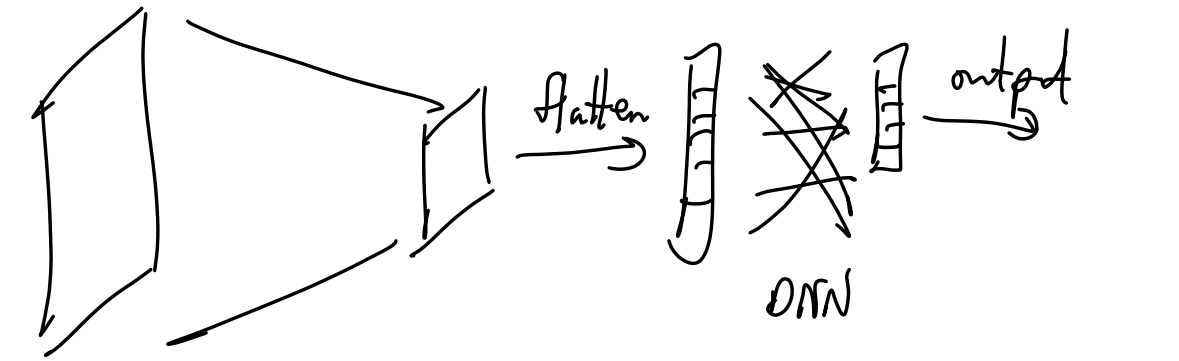
~~reduce~~ → reduce overfitting, help computational overhead

→ max(a, b, c, d) "2×2 max pooling"

avg (---) " avg "

Friday, February 12, 2021 4:40 PM

CNN is a DNN w/ massive weight sharing



Conv. layers
+ pooling

"classic CNN arch."