CS 360: Machine Learning

Prof. Sara Mathieson

Fall 2019

HAVERFORD COLLEGE

Admin

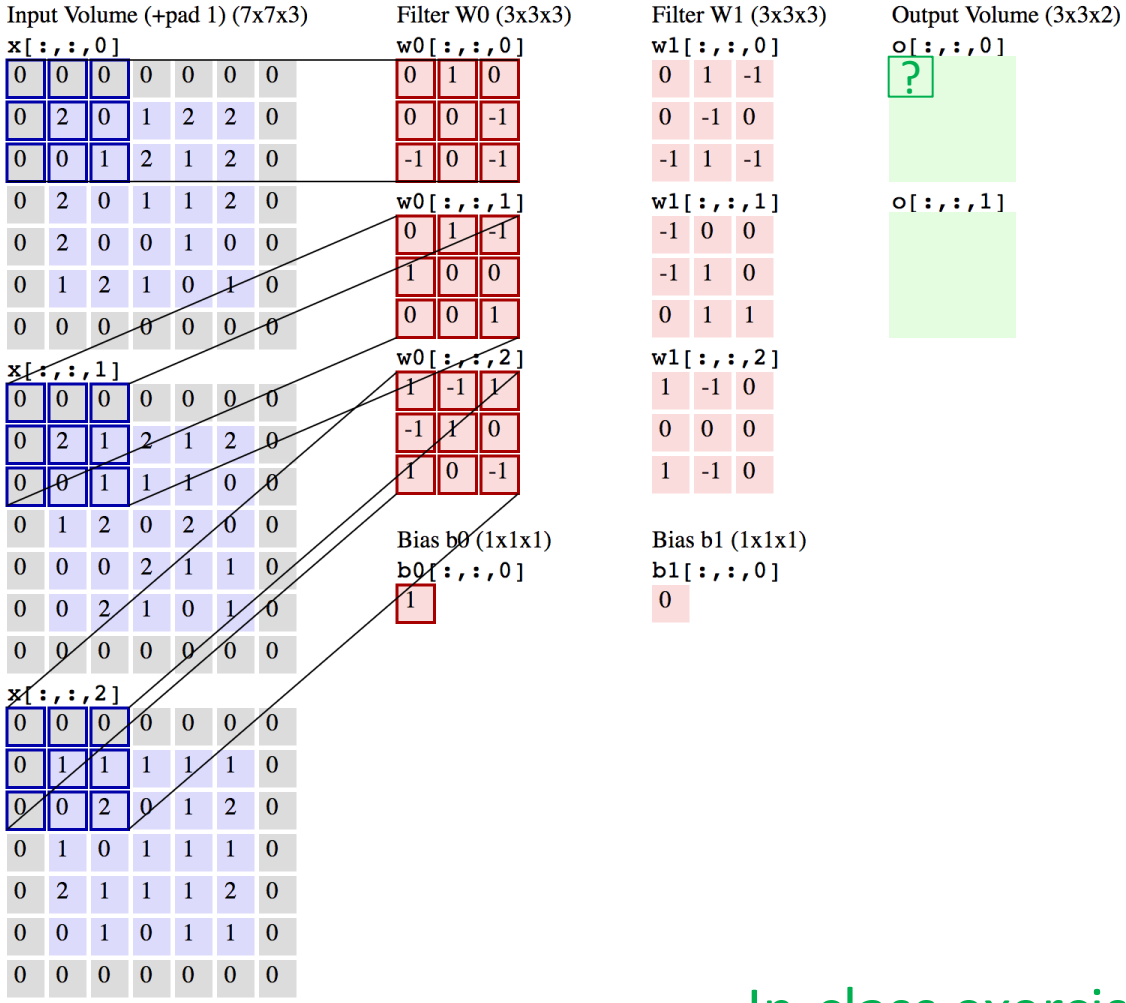
- Office hours **Today 12:30-1:30pm** (in lab)
- Office hours tomorrow **3-4pm** (in lab)
- No office hours Friday
- Reminder: vote of presentation times!
- **Midterm 2**: Thursday (in-class)
  - take home due Tues Nov 26
- Nov 28-29: Thanksgiving break!

Outline for November 19

- Recap cross-correlations
- CNN handout problems
- Ensembles and practice problems
- SVMs (next time!)

Outline for November 19

- Recap cross-correlations
- CNN handout problems
- Ensembles and practice problems
- SVMs (next time!)



In-class exercise

Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

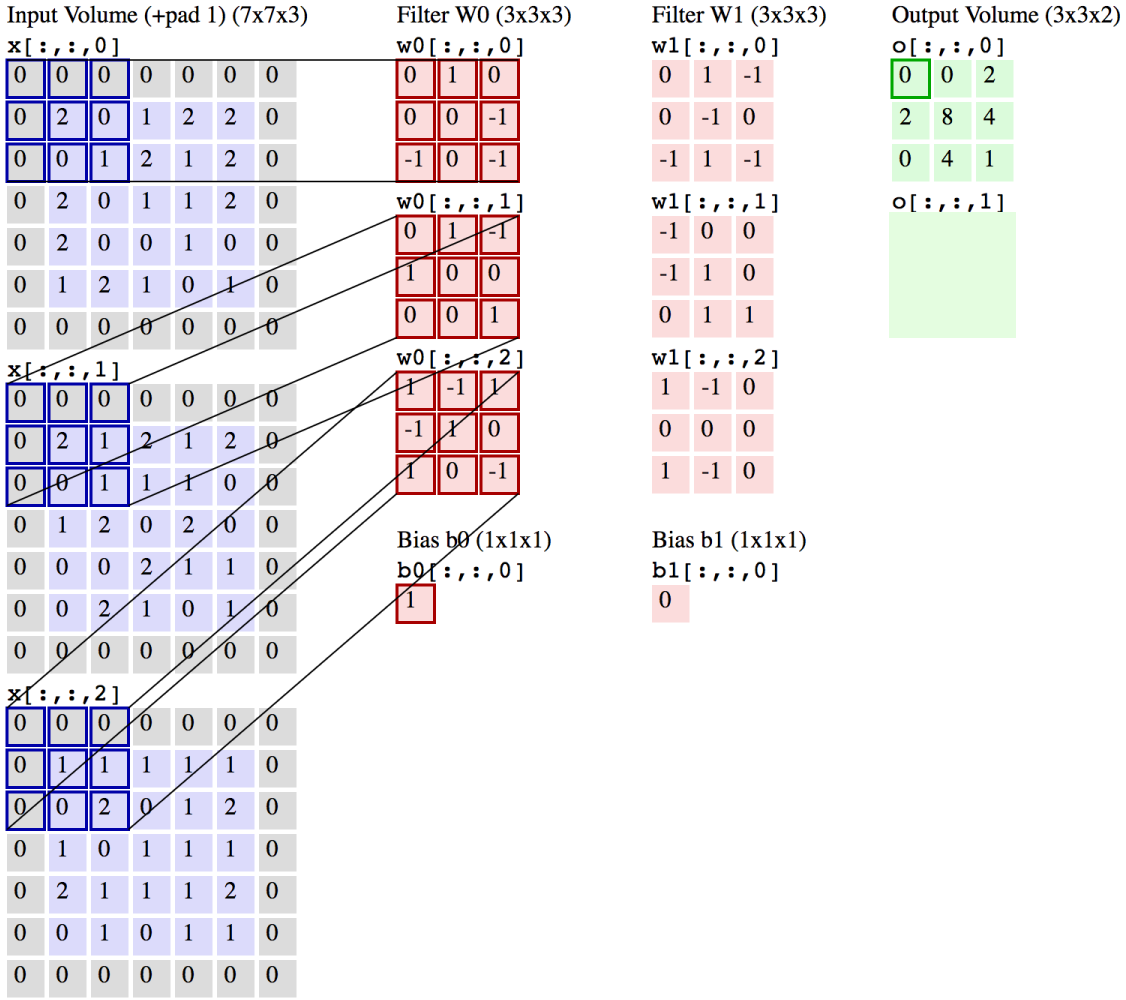


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

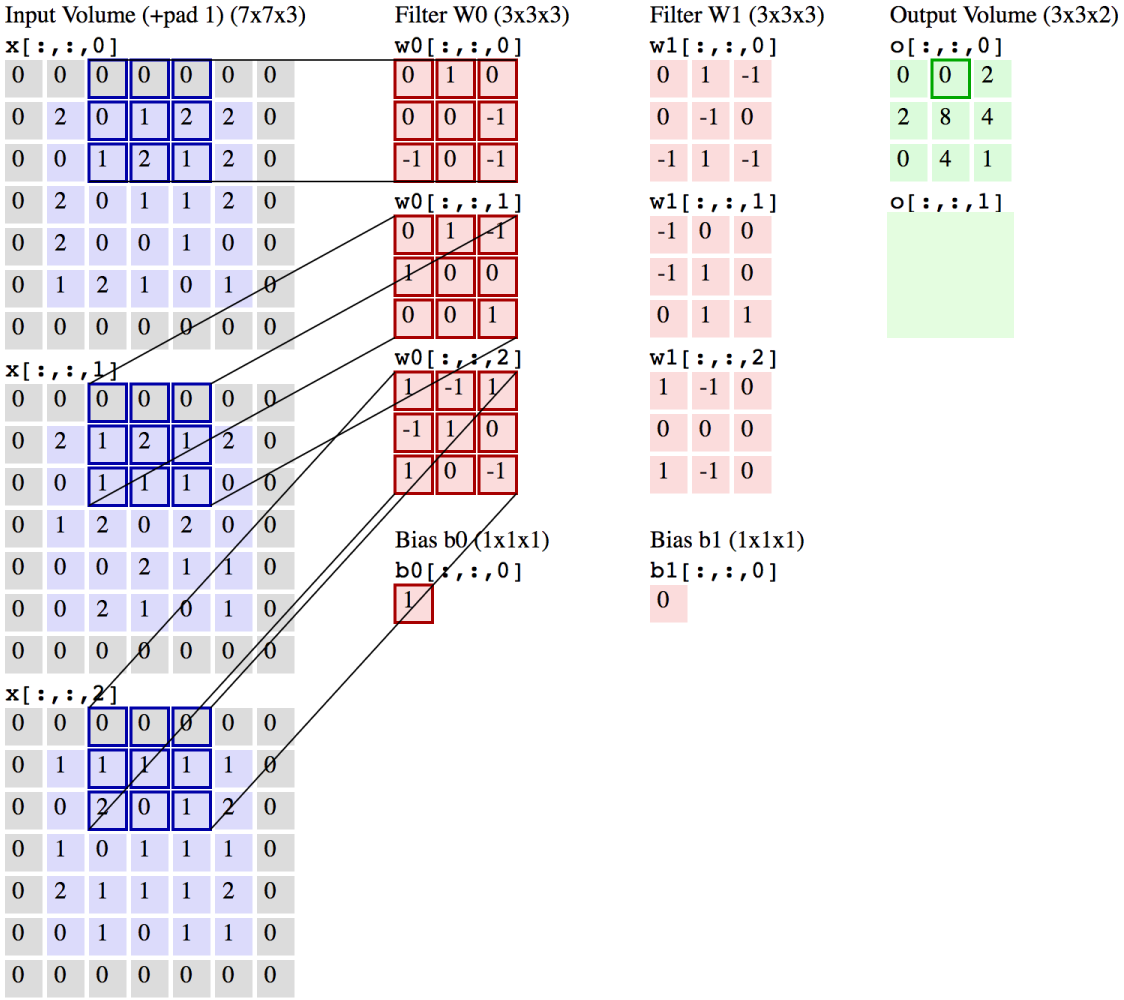


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

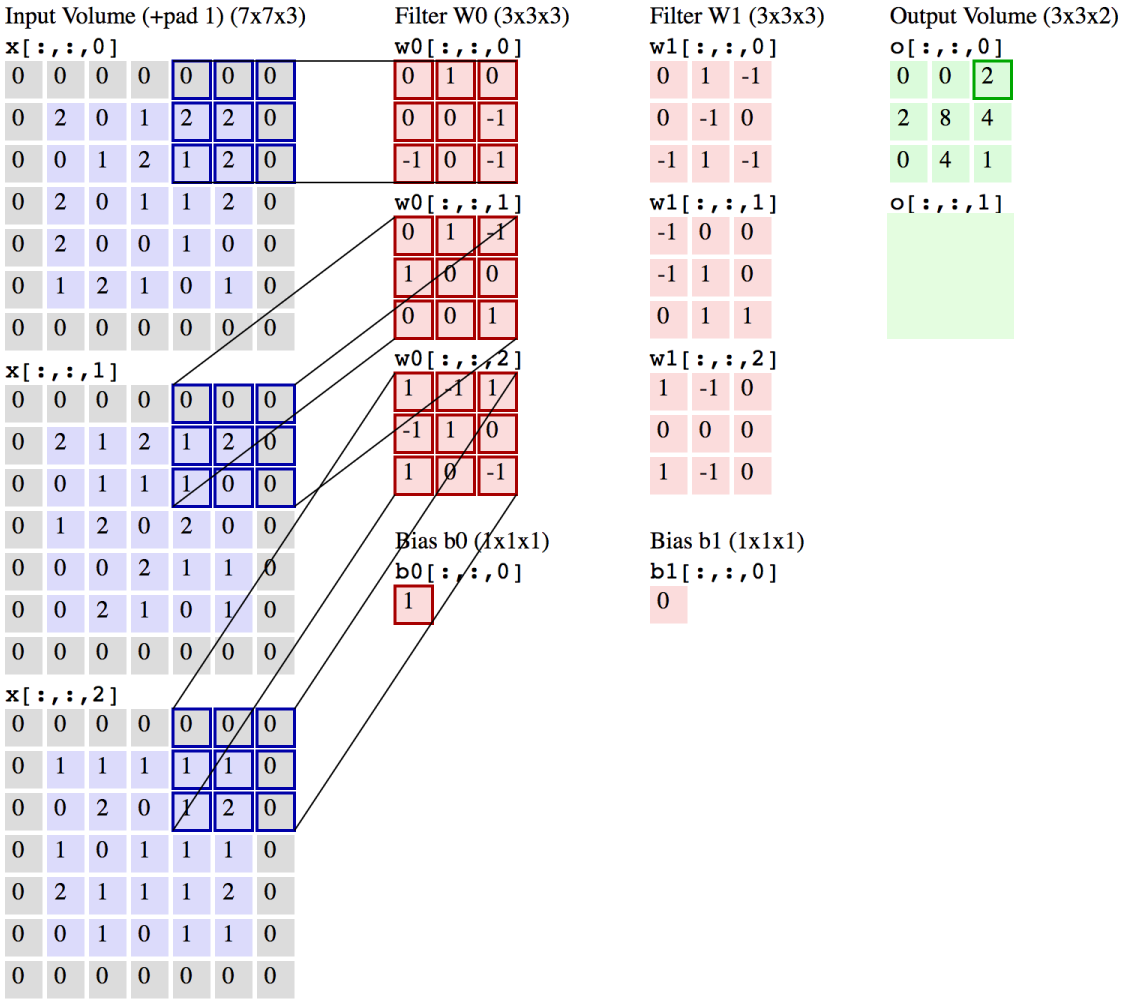


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

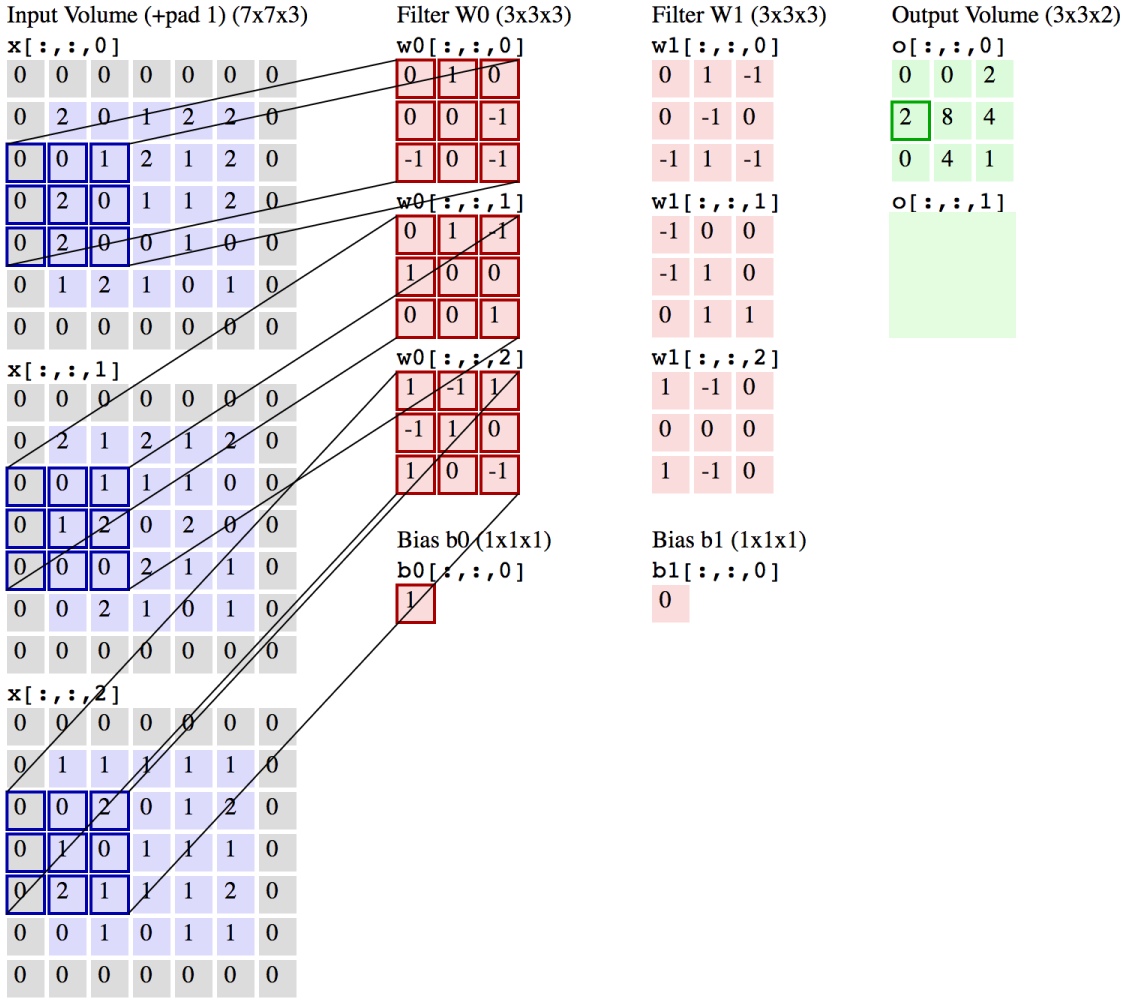


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

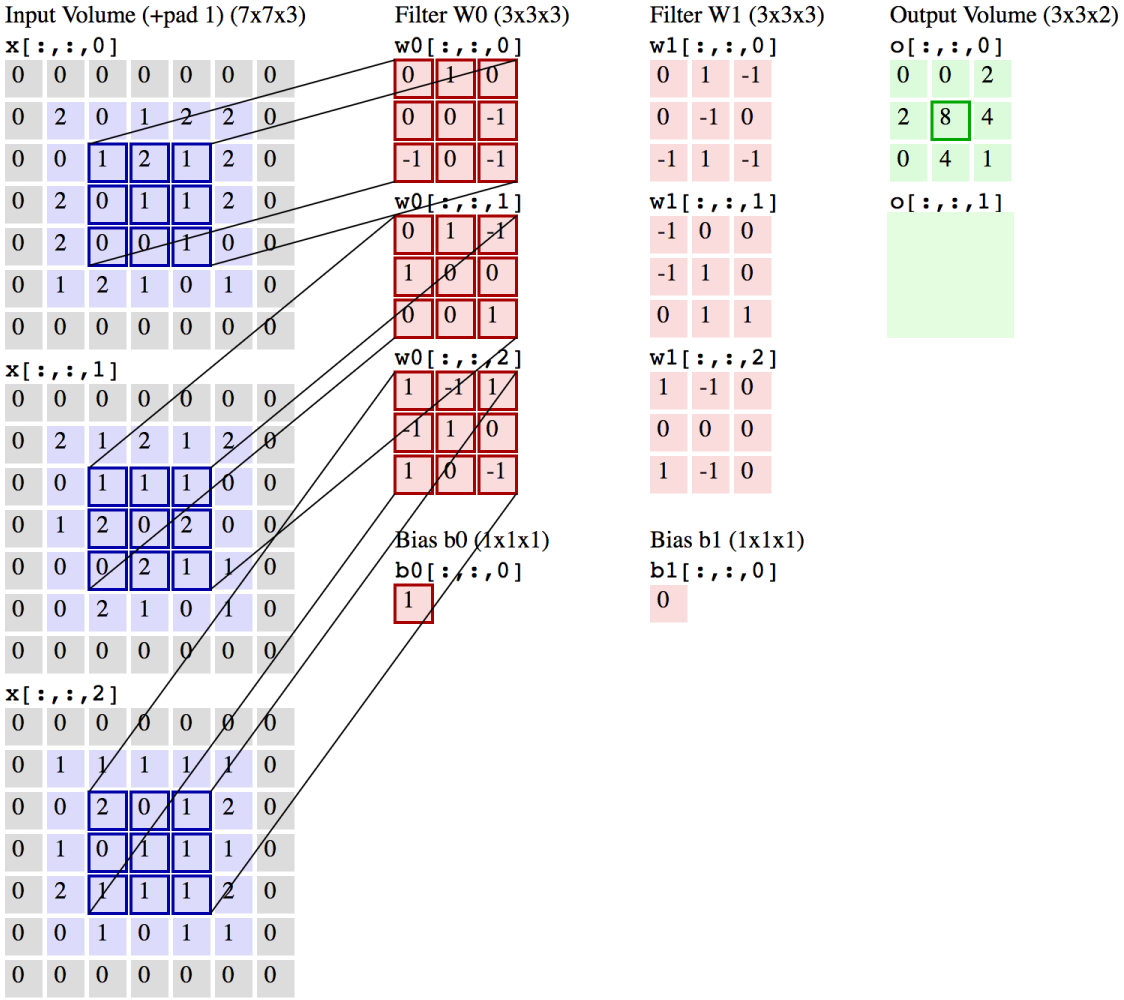


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

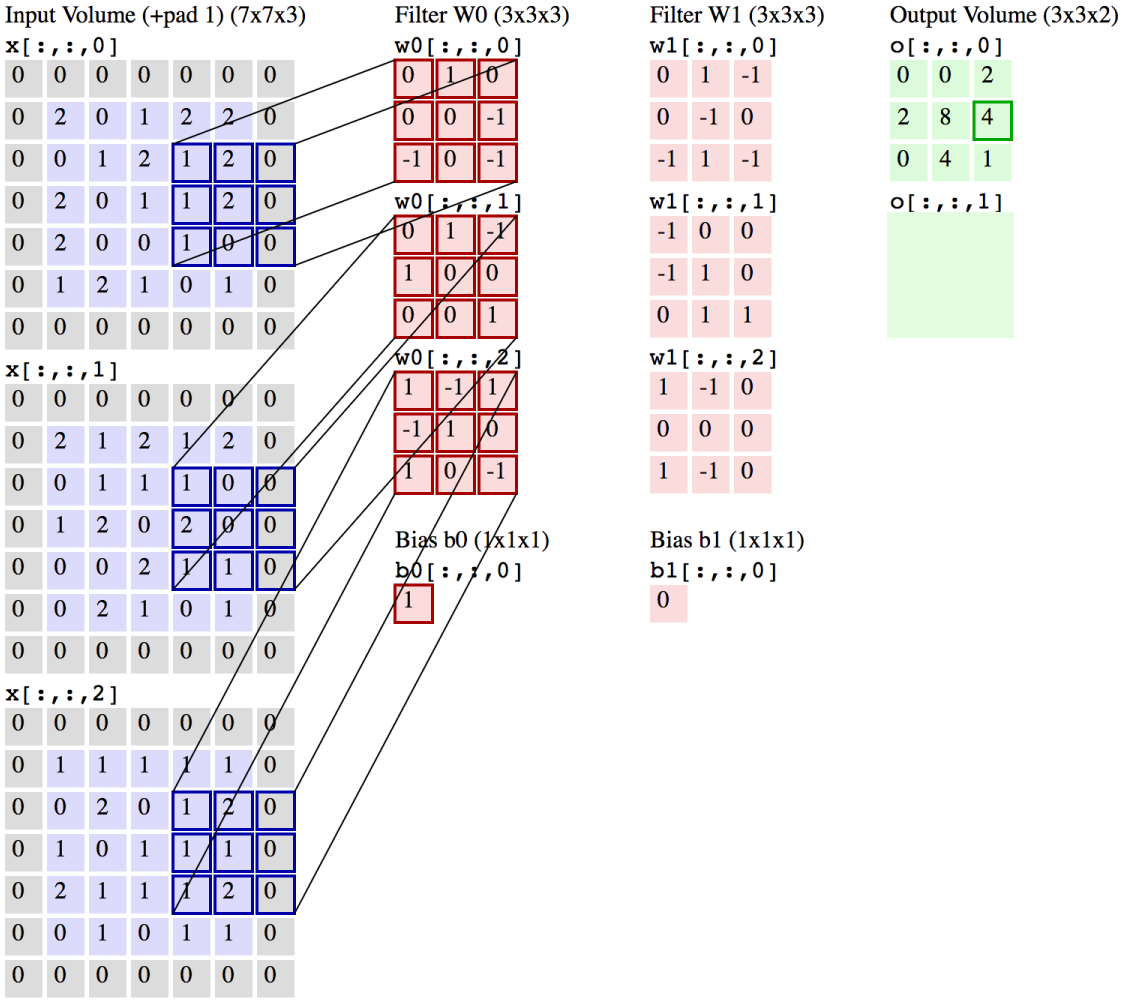


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

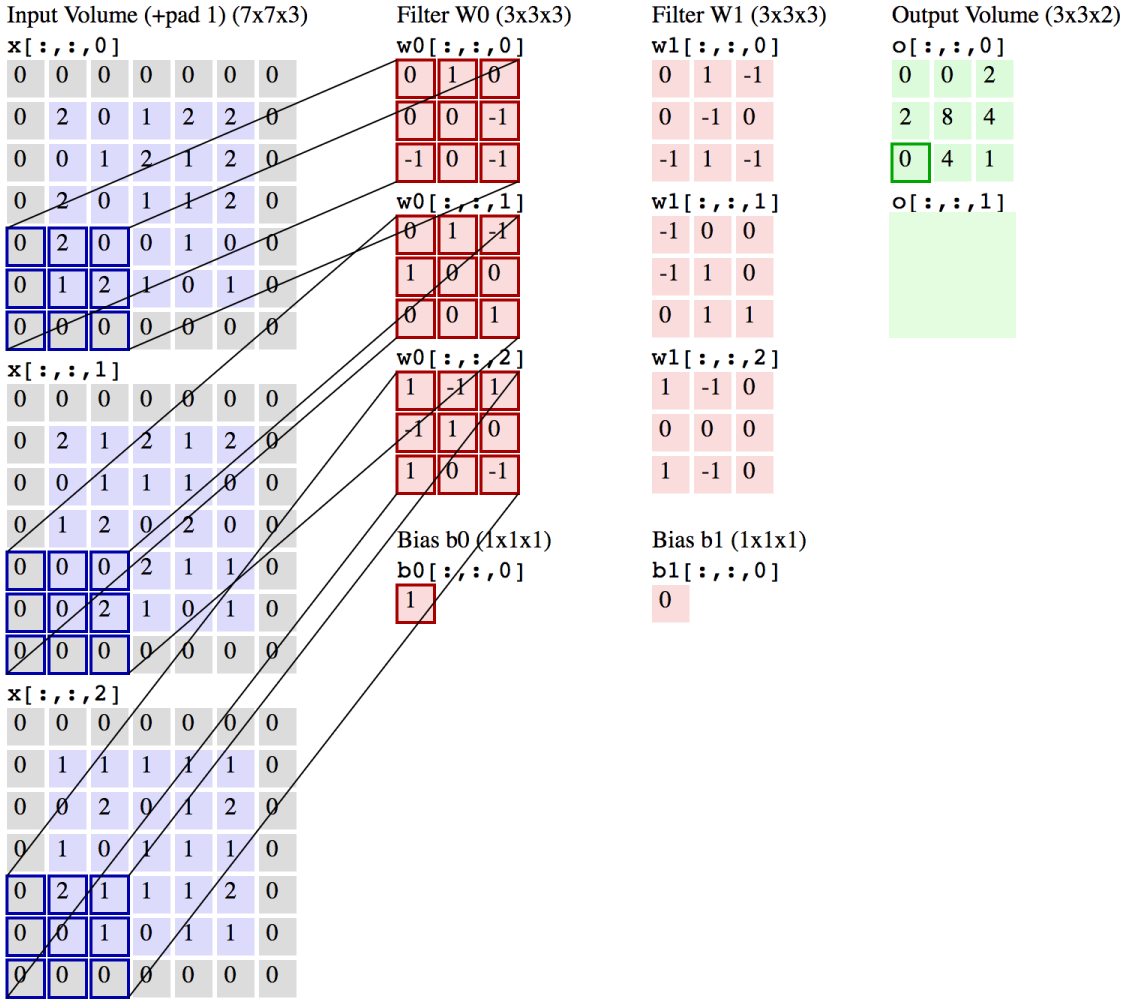


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

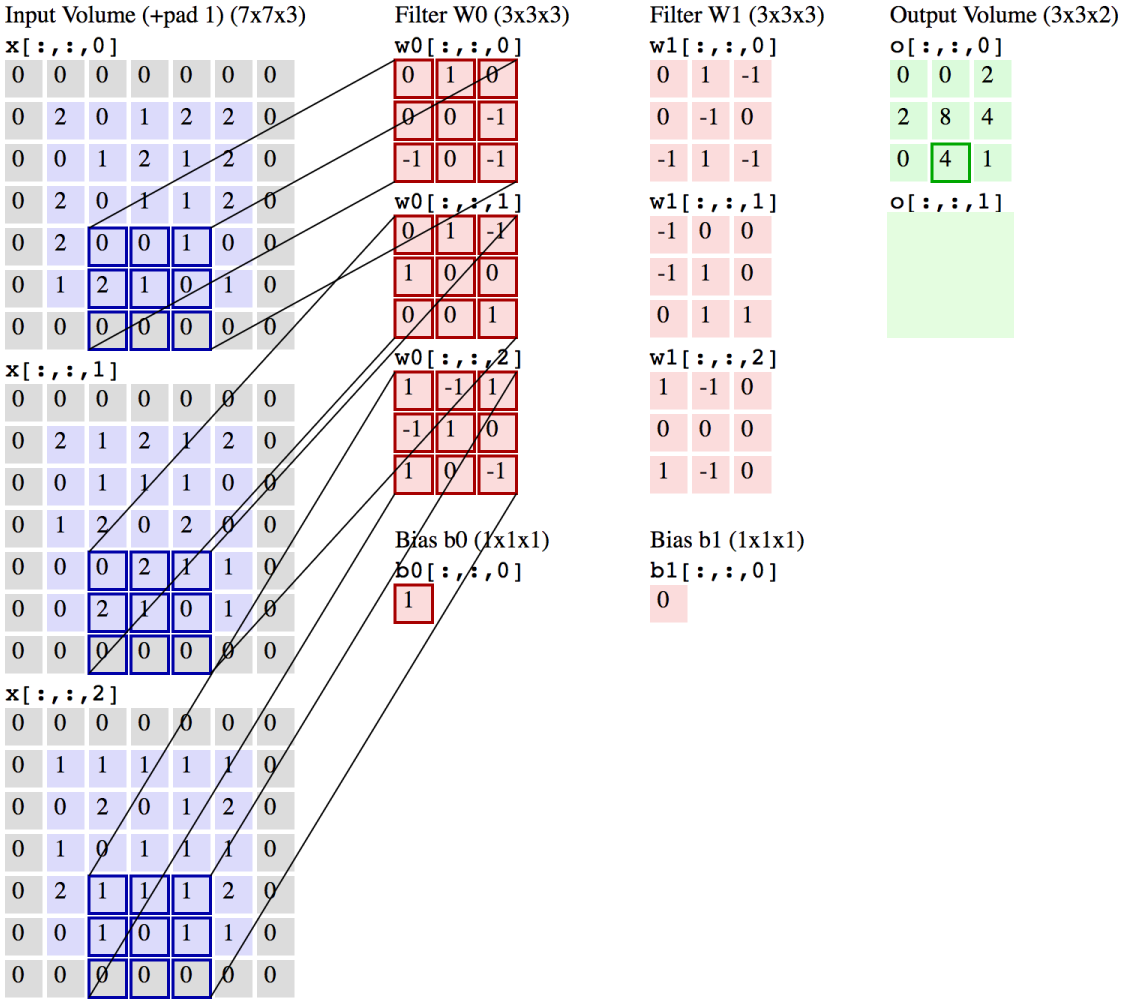


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

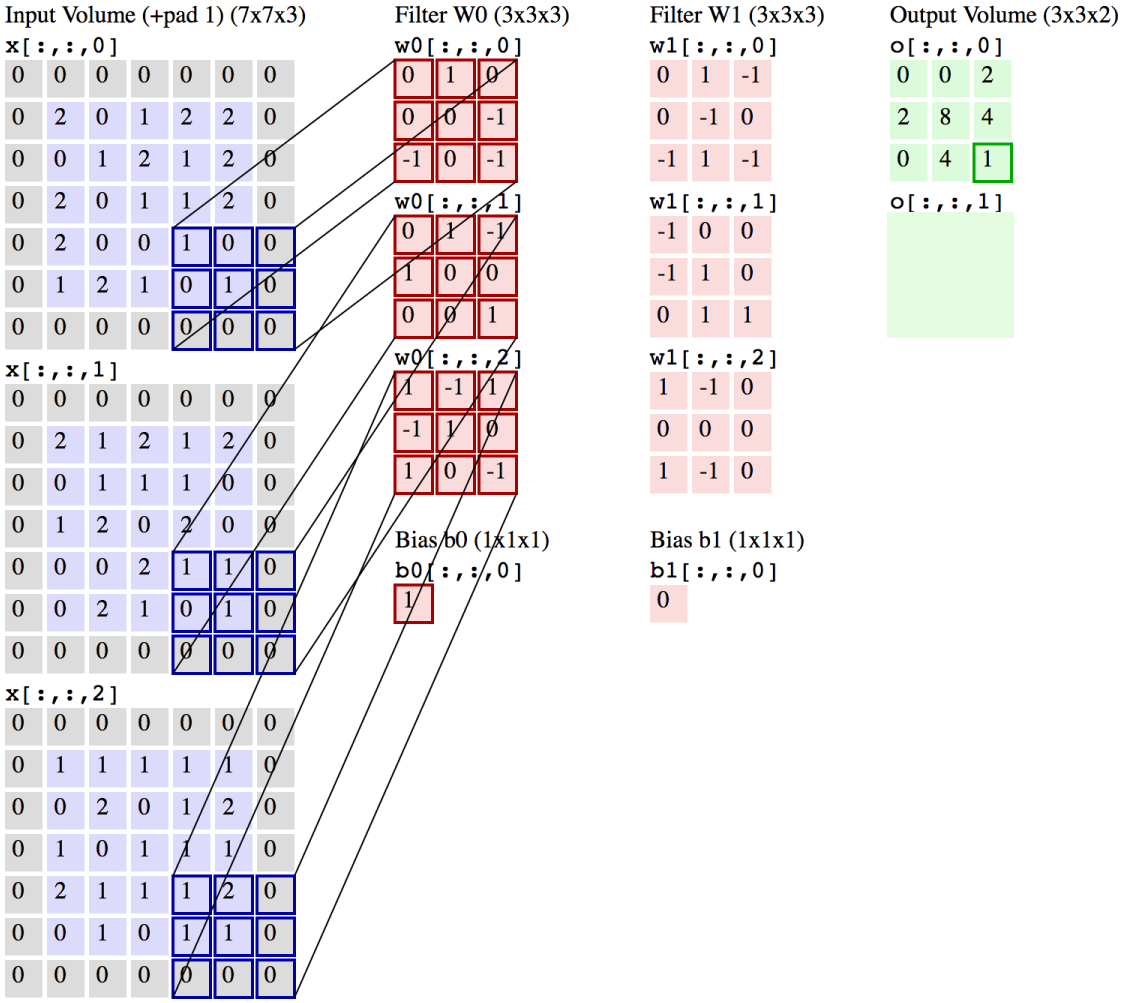


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

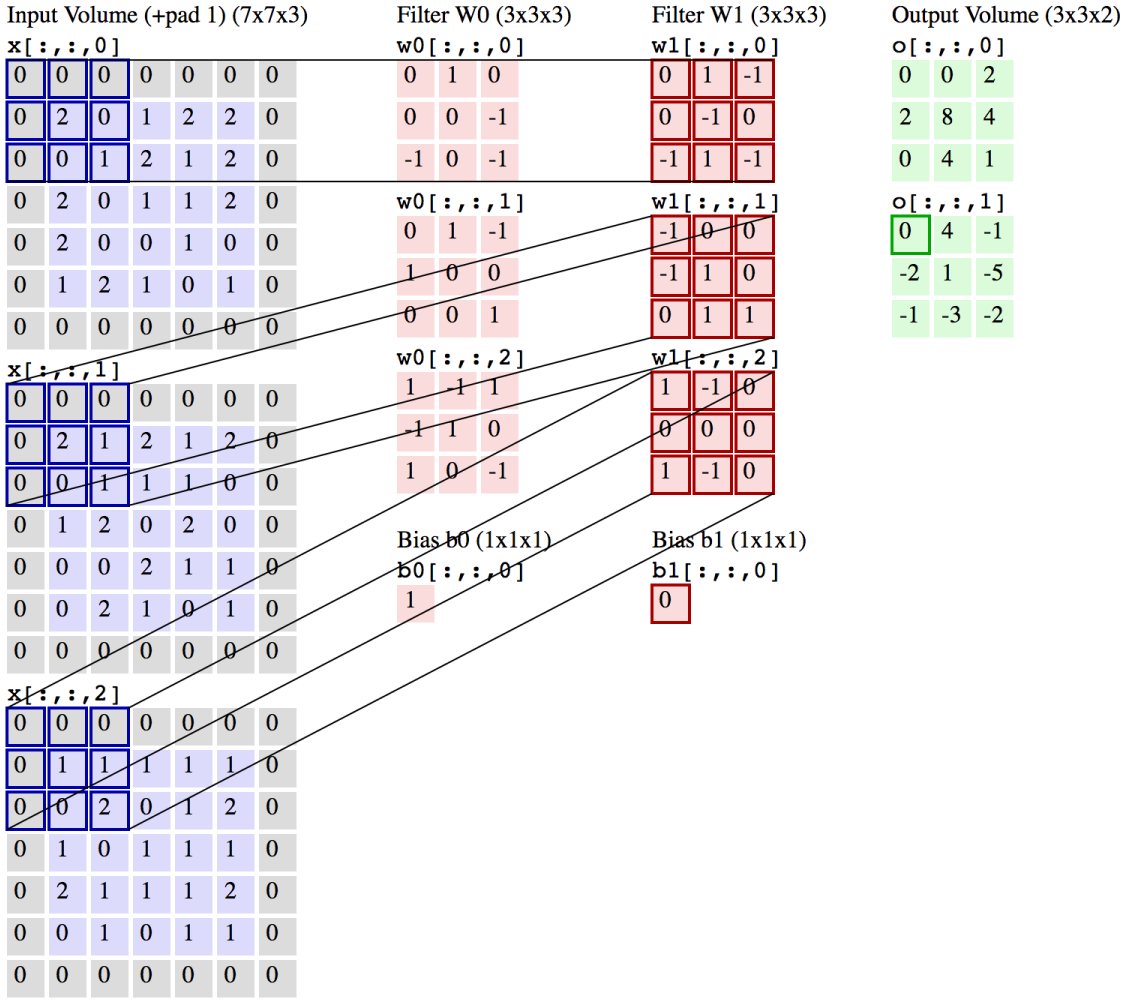


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

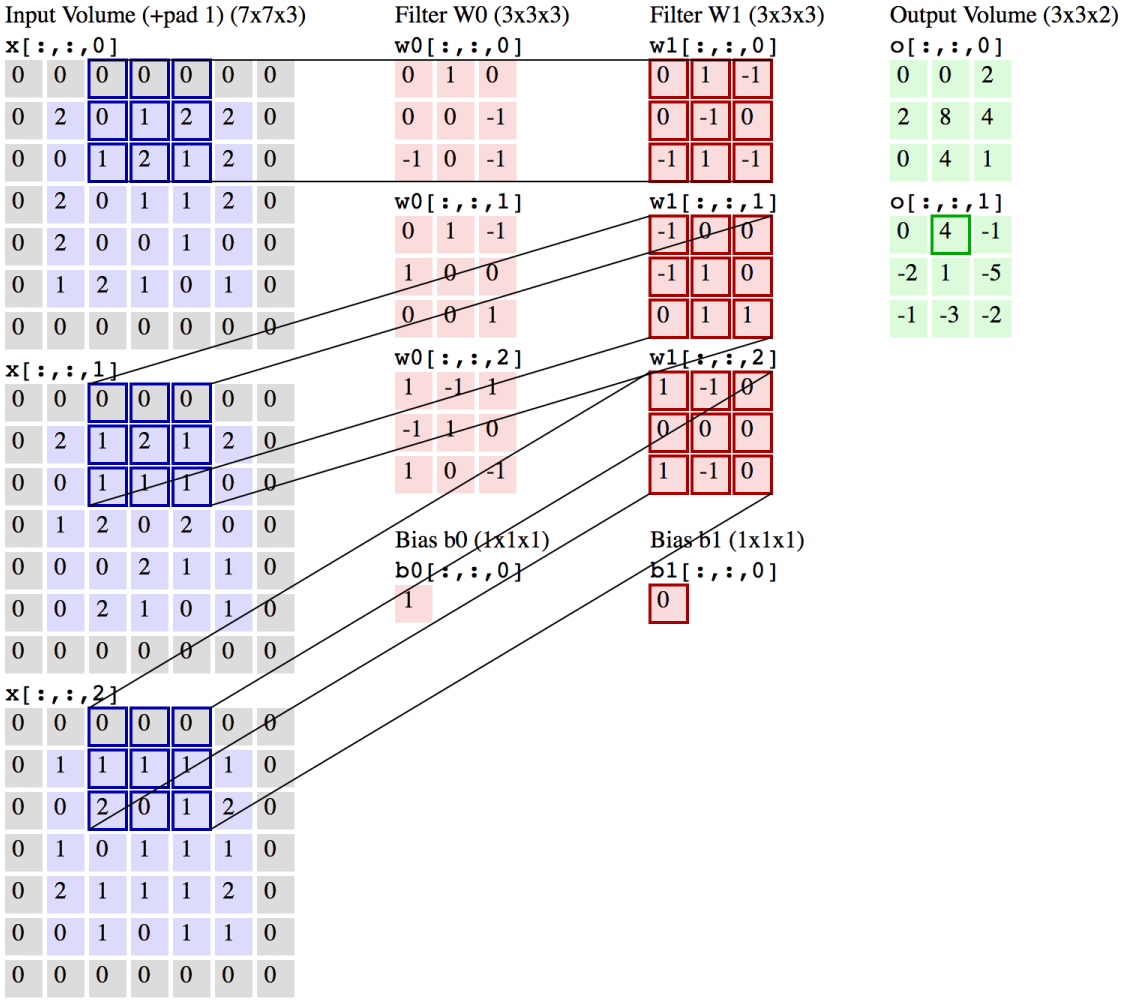


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

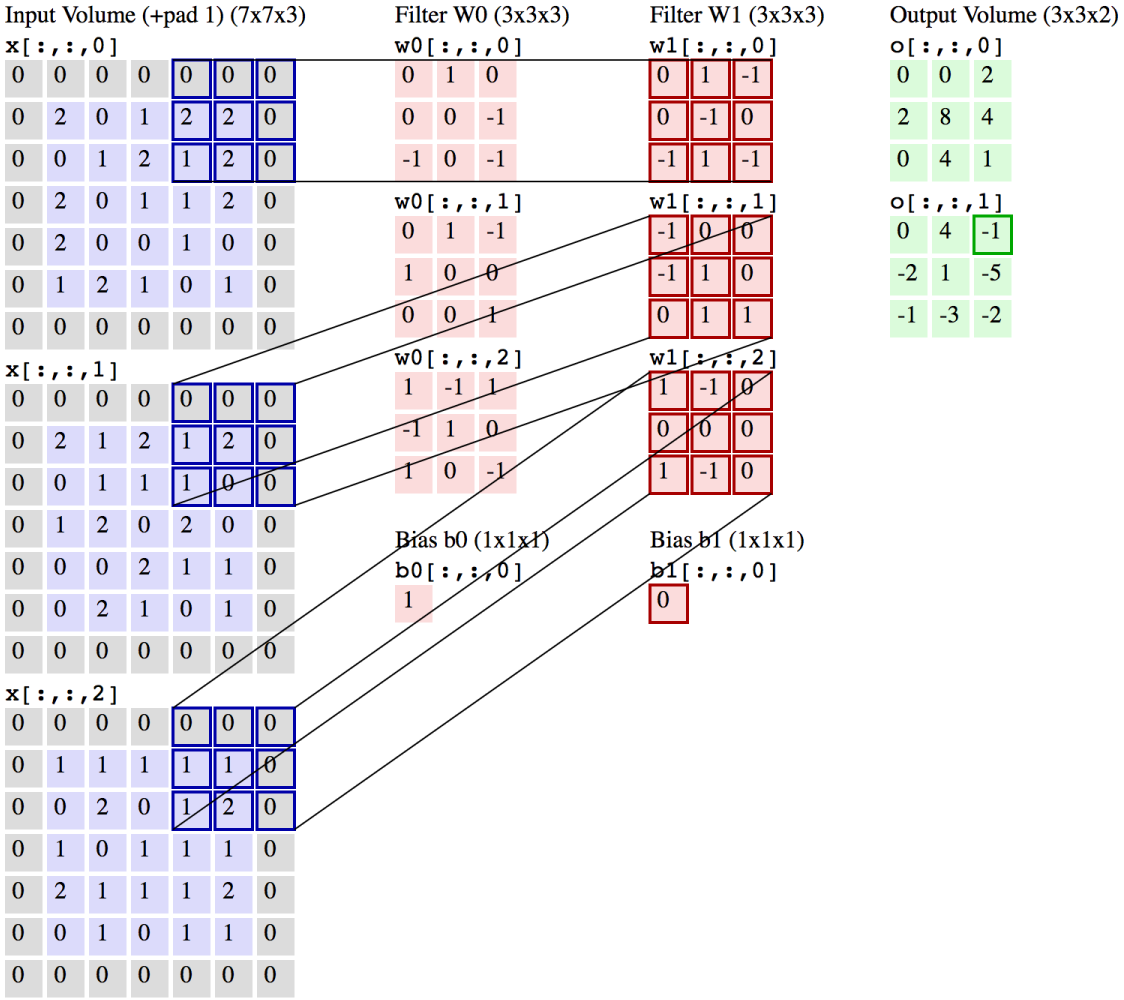


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

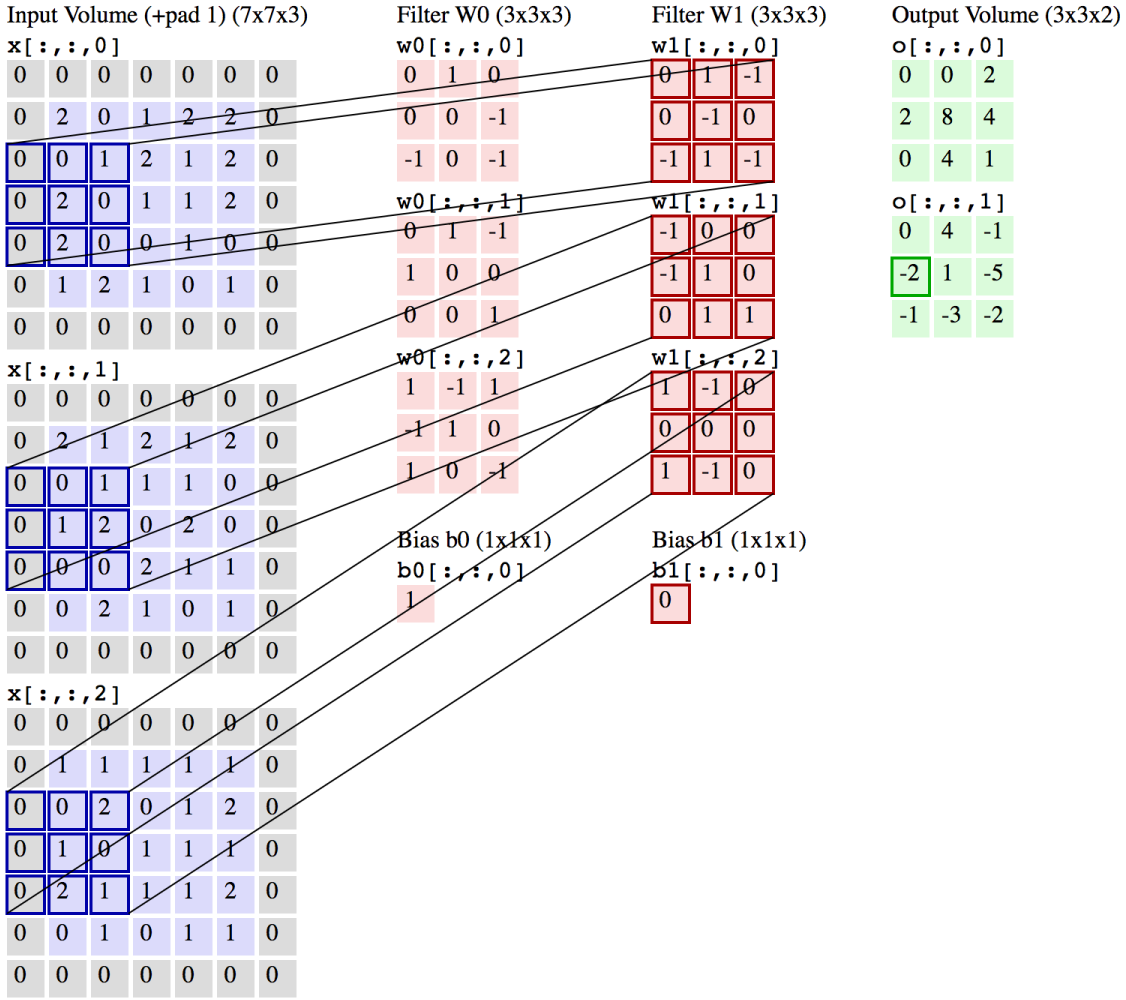


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

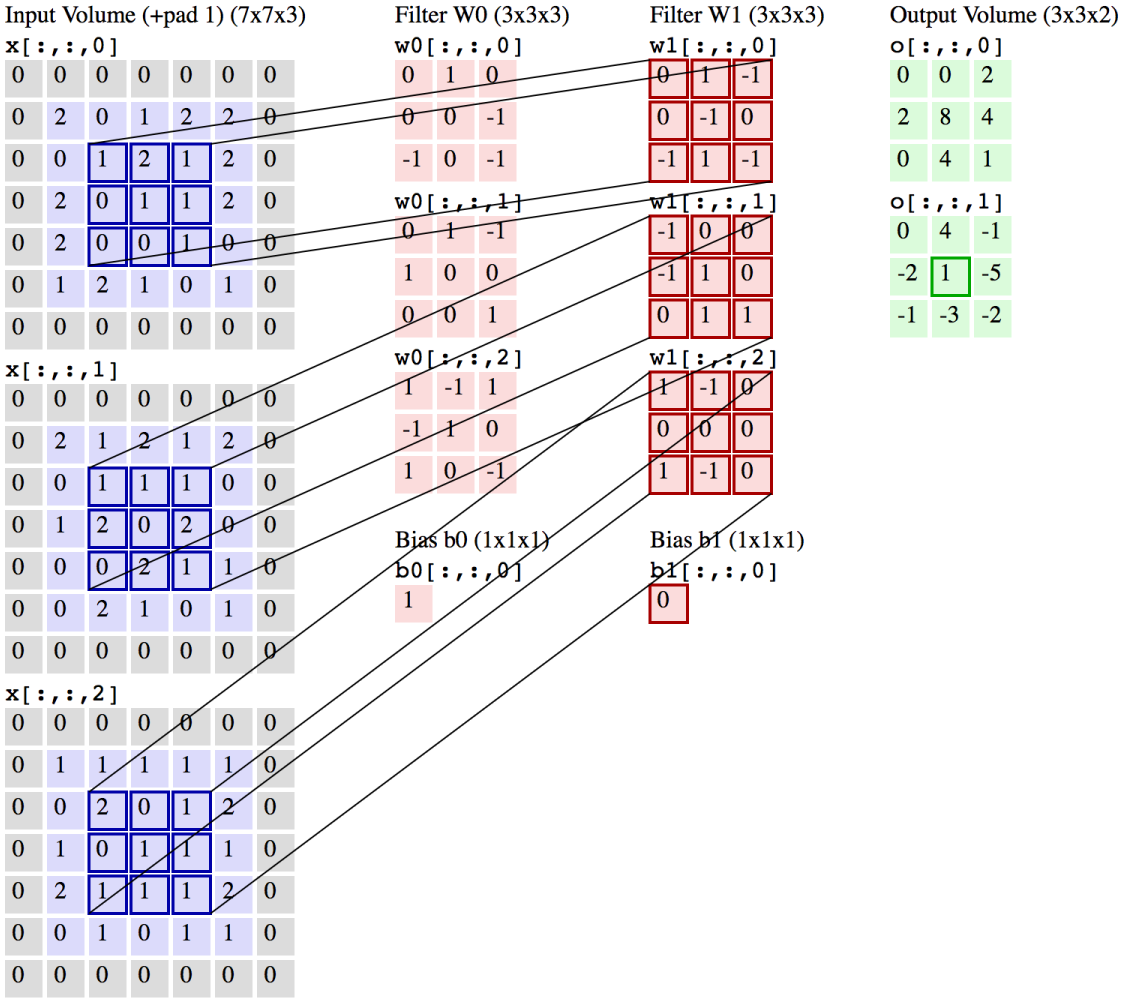


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

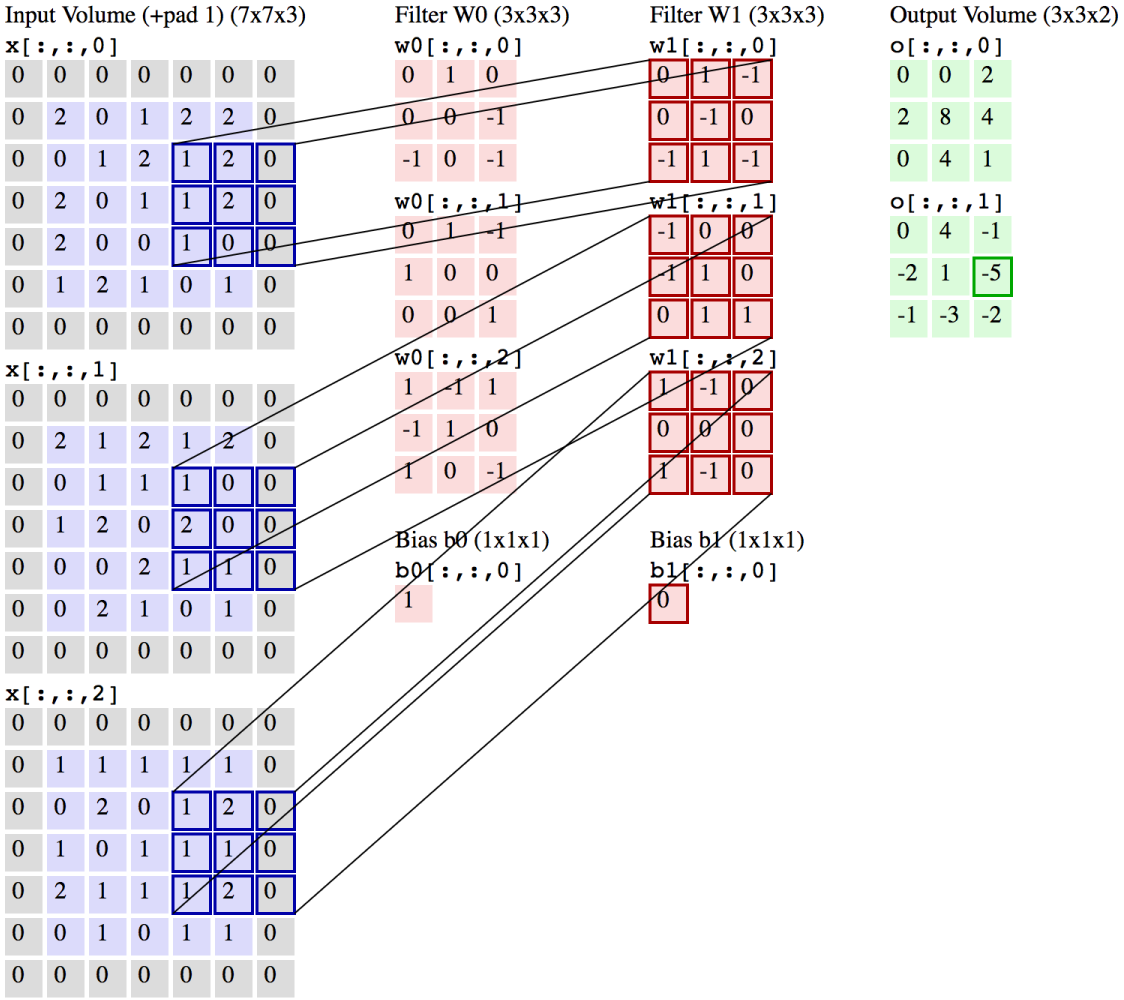


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

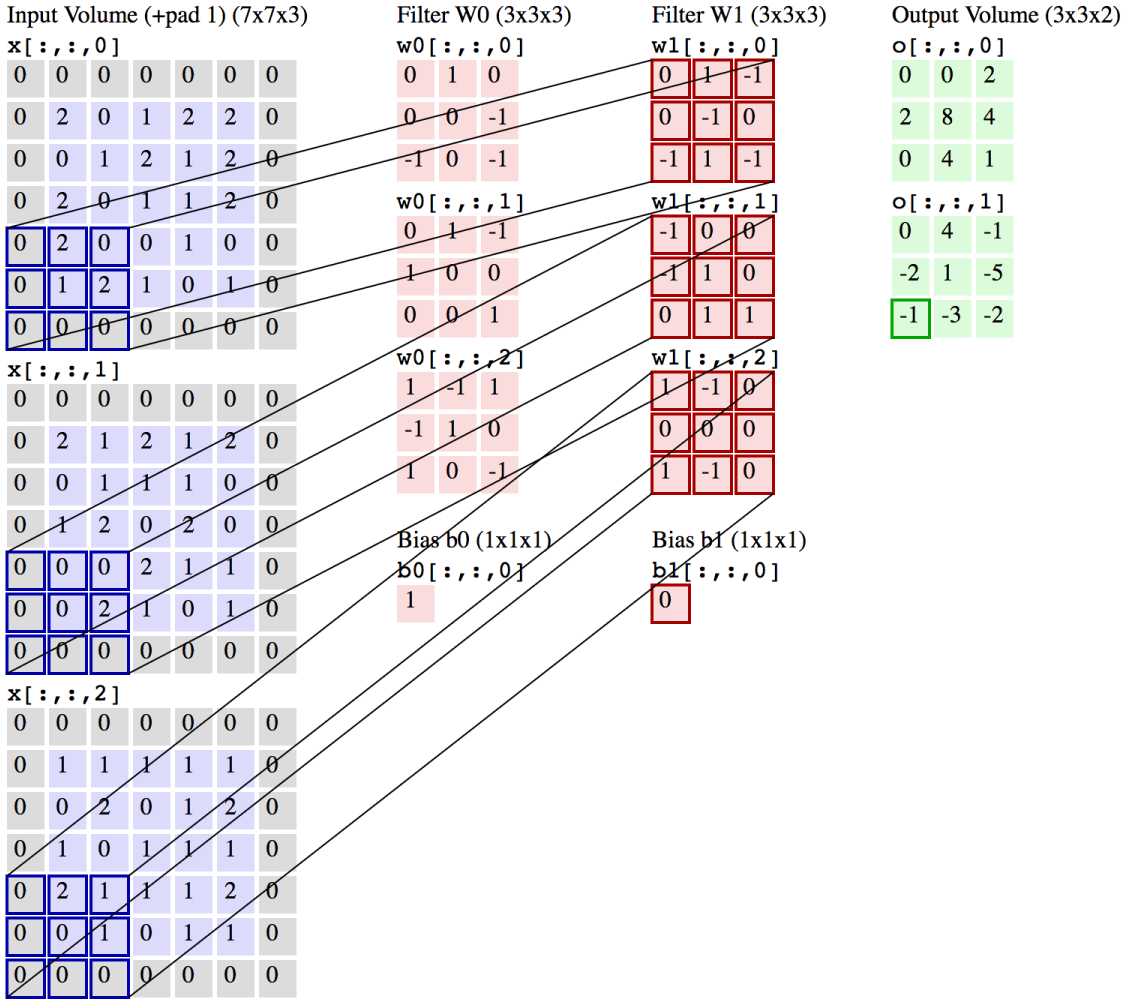


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

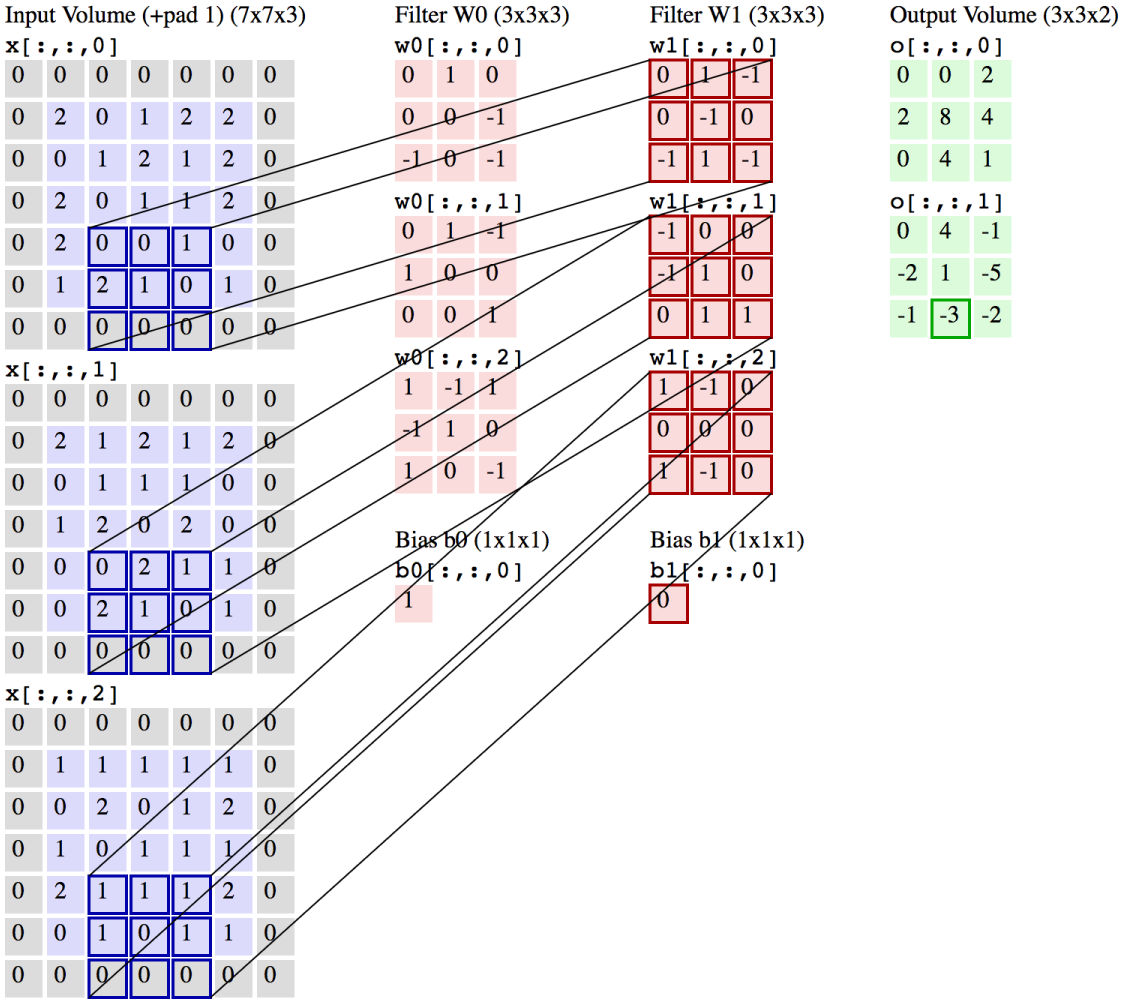


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

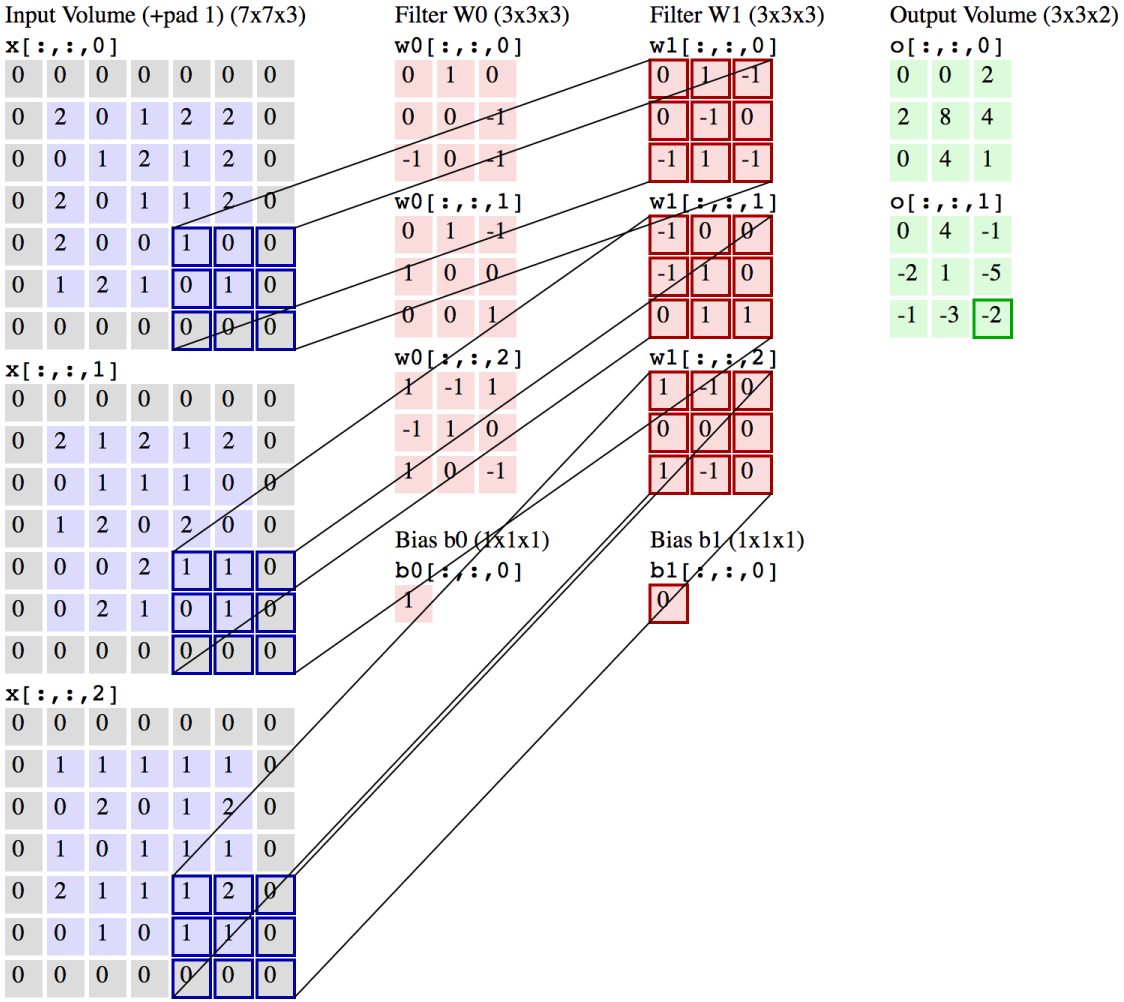
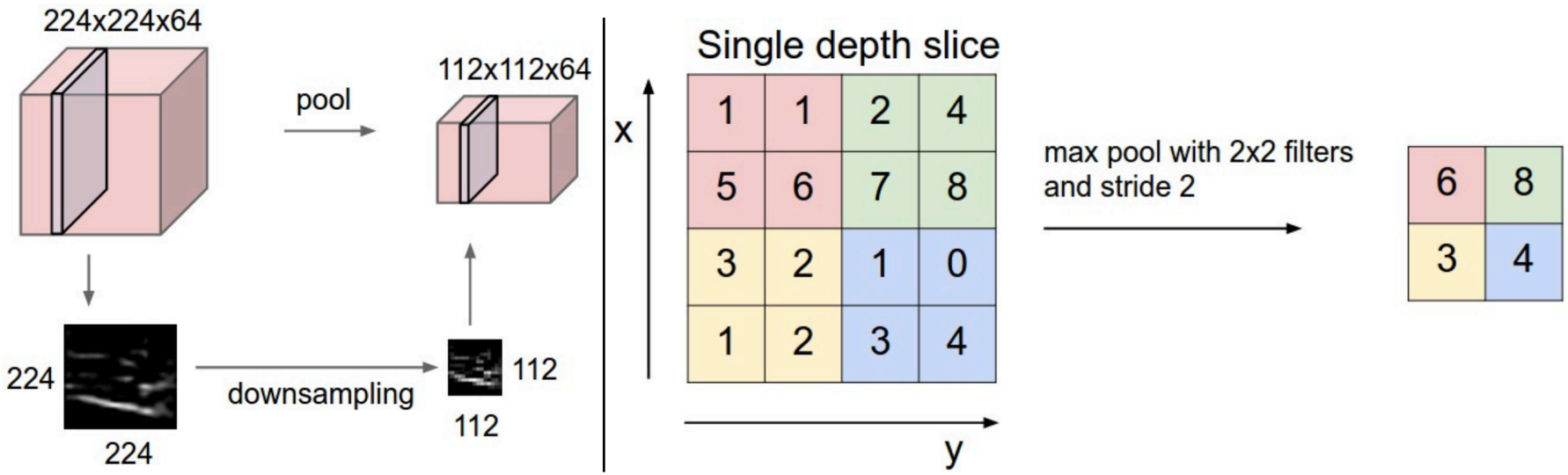


Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

Outline for November 19

- Recap cross-correlations
- CNN handout problems
- Ensembles and practice problems
- SVMs (next time!)

Pooling



Pooling layer downsamples the volume spatially, independently in each depth slice of the input volume. **Left:** In this example, the input volume of size [224x224x64] is pooled with filter size 2, stride 2 into output volume of size [112x112x64]. Notice that the volume depth is preserved. **Right:** The most common downsampling operation is max, giving rise to **max pooling**, here shown with a stride of 2. That is, each max is taken over 4 numbers (little 2x2 square).

Image: Stanford Course CS231n: http://cs231n.github.io/convolutional-networks/

Handout 18, Q4

(a) Which steps require parameter learning? (out of CONV, RELU, POOL, FLATTEN, FC)

CONV, FC

(b) First layer params 5*5*3*20 + 20 = 1520

(c) Second layer params 3*3*20*10 + 10 = 1810

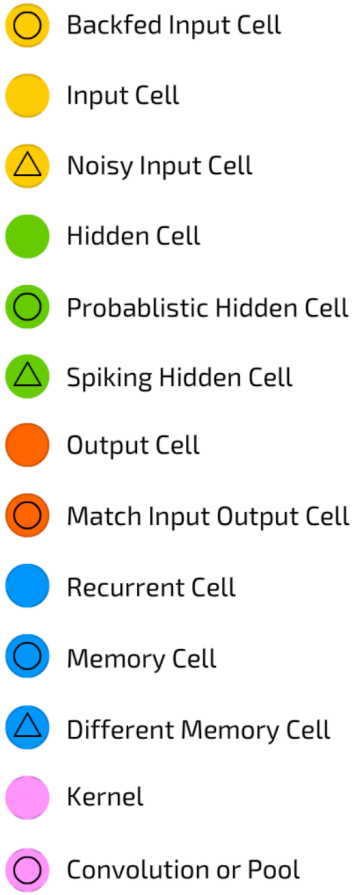
(d) Third layer params 8*8*10*10 + 10 = 6410

(e) Total # params 9740

If we had a FC with p_1=100 and p_2=50, we would have 312,860 params to learn (check this after class). CNN is much better!

A mostly complete chart of

Neural Networks



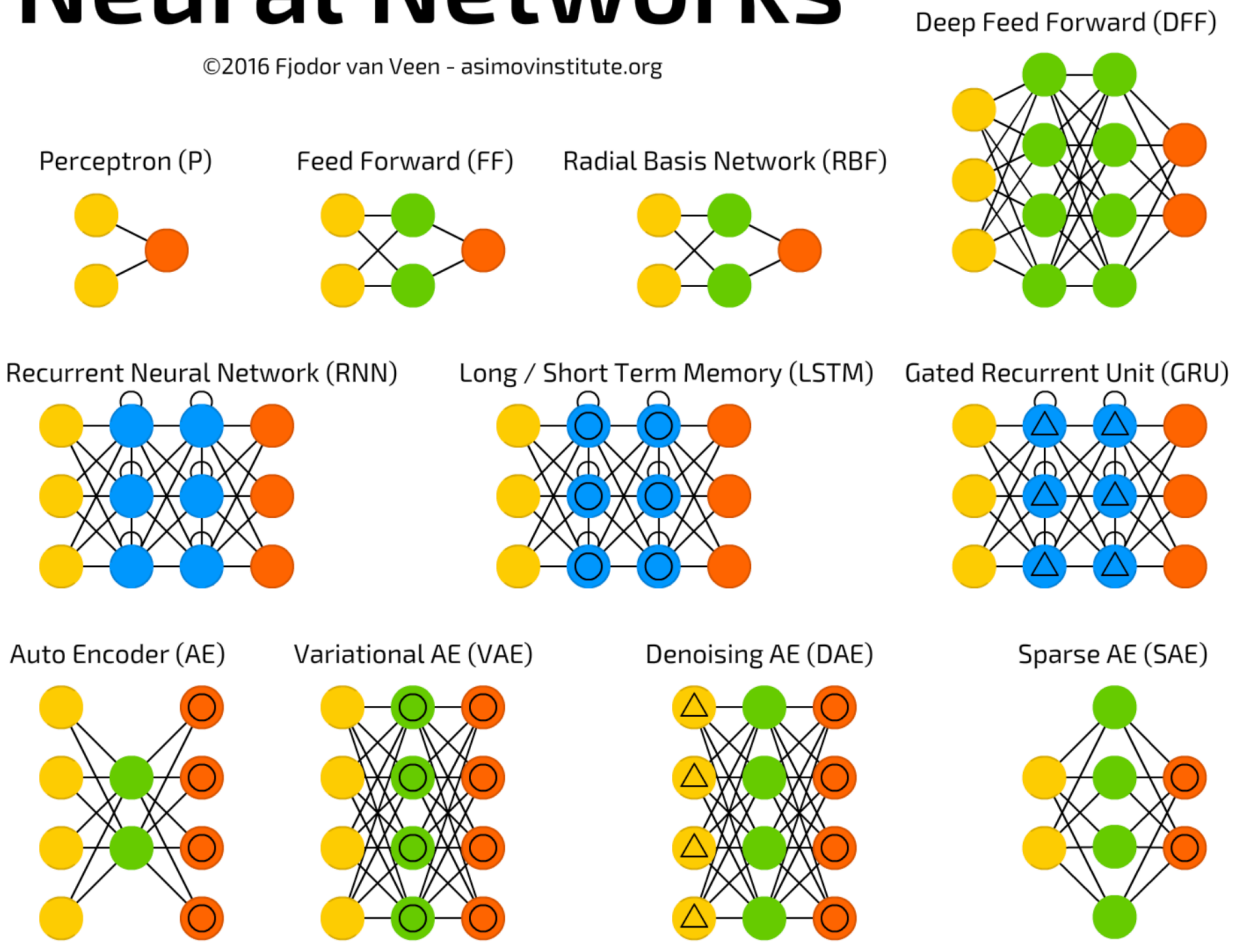


Image from Neural Network Zoo: http://www.asimovinstitute.org/neural-network-zoo/

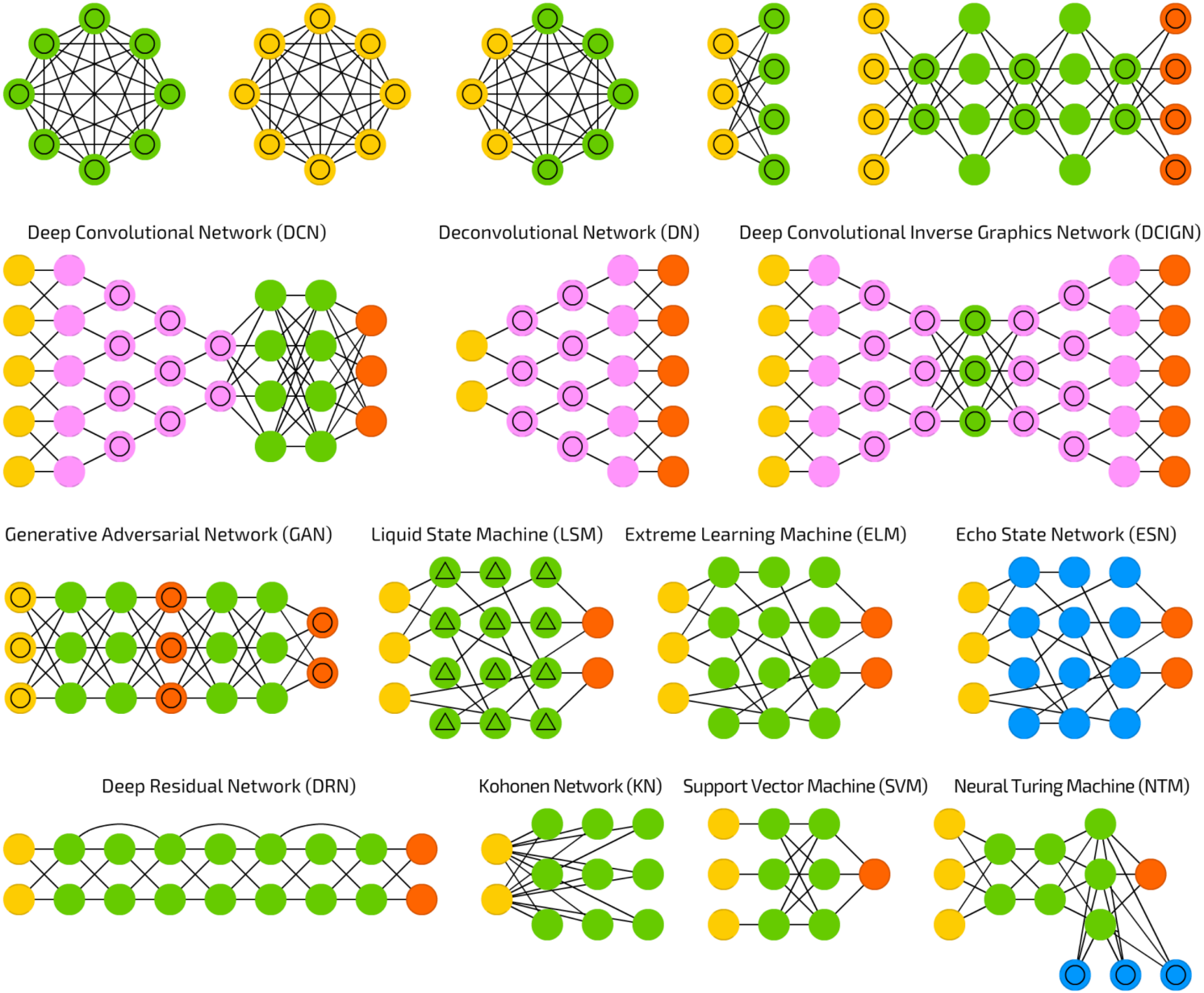
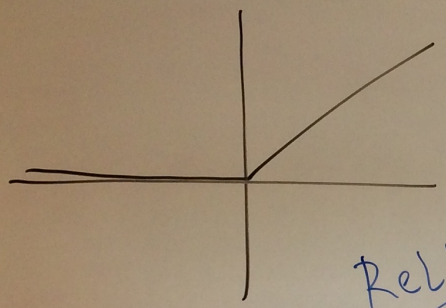


Image from Neural Network Zoo: http://www.asimovinstitute.org/neural-network-zoo/



ReLU

$f(x) = \max(0, x)$

softmax

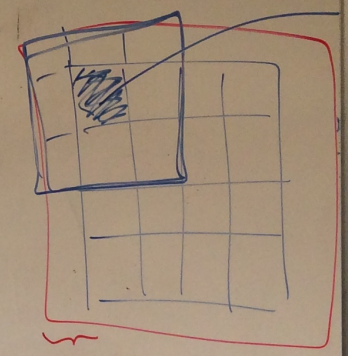
$$\boxed{\frac{e^{s_k}}{\sum_{j=1}^{K} e^{s_j}} = p(\text{class } k)}$$

$\Rightarrow$ prob. dist.

$\hat{y} = 2$

one-hot

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$



pred = 1

Cross-entropy

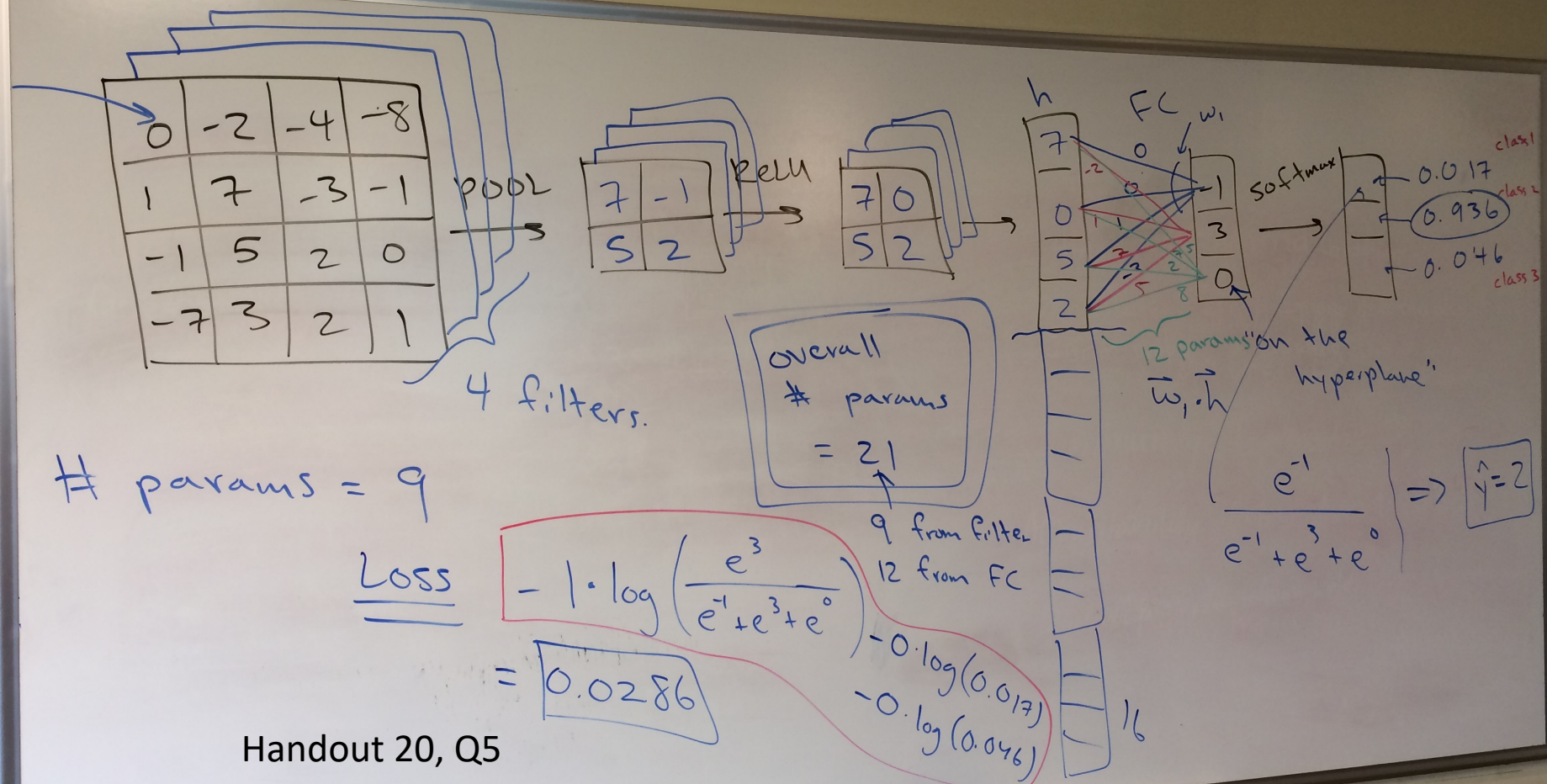
true

predicted probs.

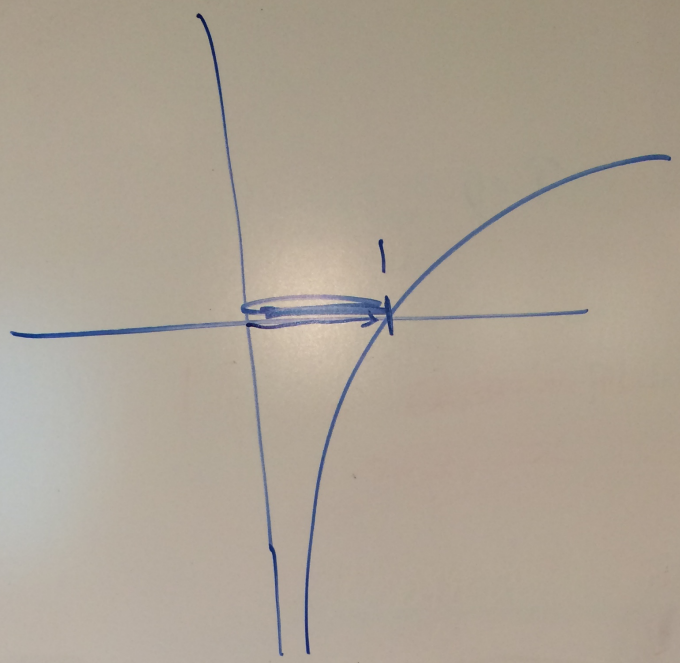
$$H(y, p(\hat{y})) = -\sum_{k=1}^{K} y_k \log(p_k(\hat{y})) = \boxed{-\log(p(y))}$$

true

Handout 20, Q5



Handout 20, Q5



Why use log in the loss function?

If the probability of the “right” label is close to 1, we add very little to the loss function. If this probability is small, we add a high number to the loss function.

Outline for November 19

- Recap cross-correlations
- CNN handout problems
- Ensembles and practice problems
- SVMs (next time!)

Handout 20, Question 1

- First compute weighted leaf labels

$$P(+ \mid \text{sun}) = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{8} + \frac{1}{8}} = \frac{4}{7} \geq 0.5 \quad \Rightarrow +$$

Handout 20, Question 1

- First compute weighted leaf labels

$$P(+ \mid \text{sun}) = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{8} + \frac{1}{8}} = \frac{4}{7} \geq 0.5 \quad \Rightarrow +$$

$$P(+ \mid \text{rain}) = \frac{\frac{1}{12}}{\frac{1}{12} + \frac{1}{6} + \frac{1}{6}} = \frac{1}{5} < 0.5 \quad \Rightarrow -$$

Handout 20, Question 1

- Based on these labels, we can say which training points are misclassified

$$\epsilon_t = \frac{1}{8} + \frac{1}{8} + \frac{1}{12} = \frac{1}{3}$$

Handout 20, Question 1

- Based on these labels, we can say which training points are misclassified

$$\epsilon_t = \frac{1}{8} + \frac{1}{8} + \frac{1}{12} = \frac{1}{3}$$

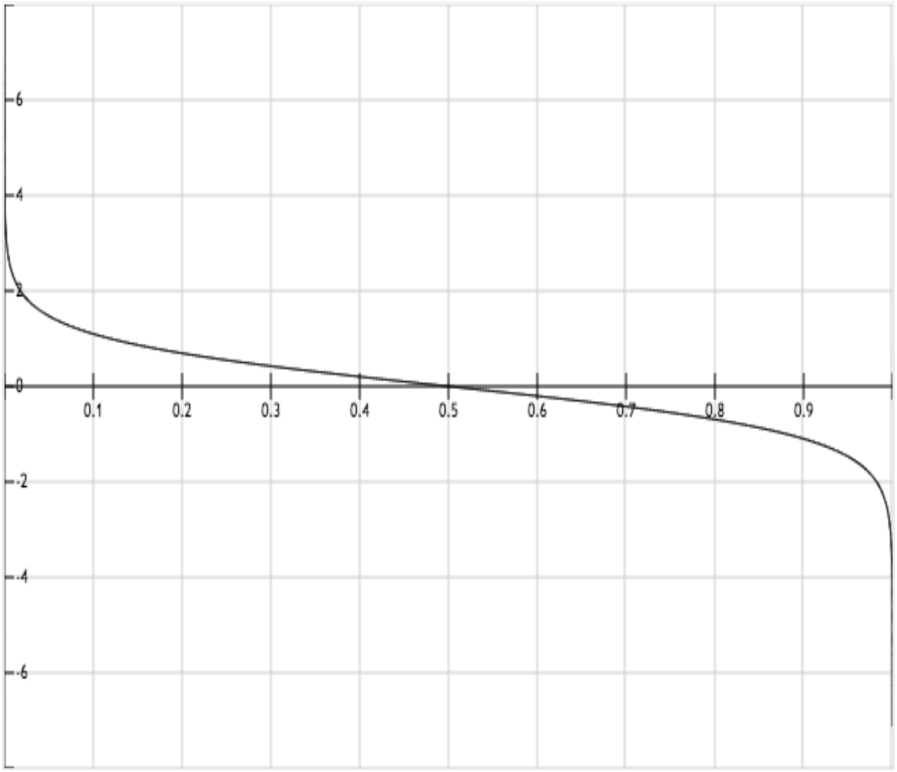
- Note if this was > 0.5, we should have chosen different leaf labels! So this "flipping" step should happen automatically
  - (exception for pathological cases)

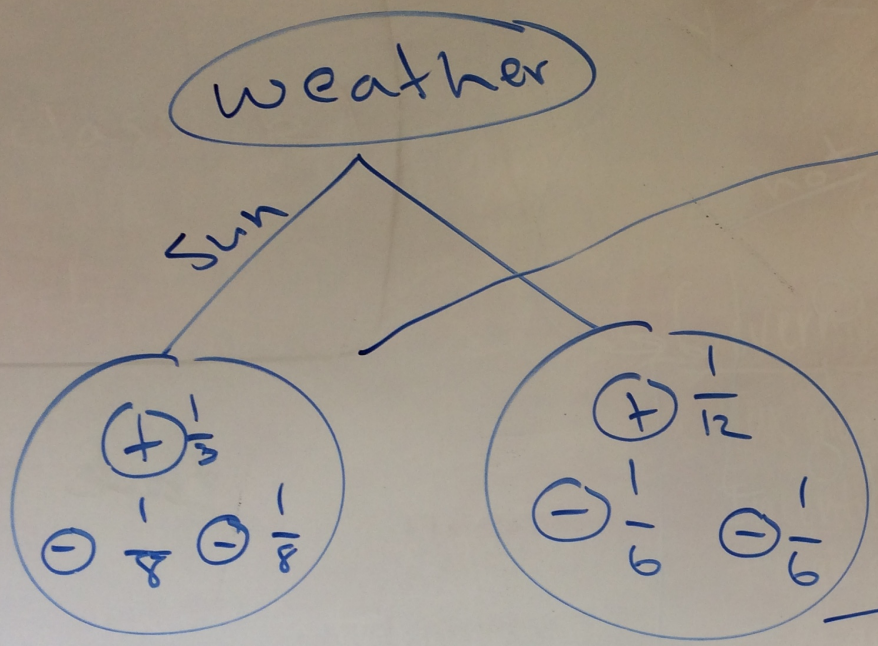
Handout 20, Question 1

- Score function: $$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$$

- Fraction: accuracy/error

- As error -> 0, score becomes high
- As error -> ½, score goes to 0





$$p(+) = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{8} + \frac{1}{8}} = \frac{4}{7}$$

$\Rightarrow$ label leaf $+$

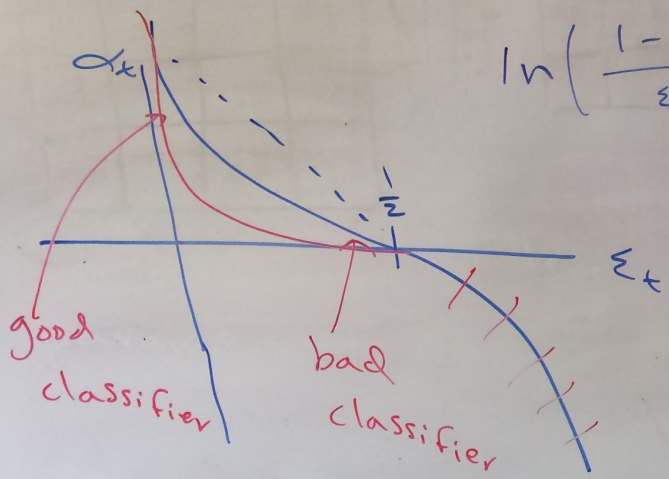
$$p(+) = \frac{\frac{1}{12}}{\frac{1}{12} + \frac{1}{6} + \frac{1}{6}} = \frac{1}{5}$$

$\Rightarrow$ label leaf $-$

Handout 20, Q1

$$\varepsilon_t = \frac{1}{8} + \frac{1}{8} + \frac{1}{12} = \frac{1}{3}$$

$$\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$$



Handout 20, Q2

Handout 20, Question 2

- r = 1/3, probability of one classifier being wrong
- T = 5, number of classifiers
- R = number of votes for the wrong class
- If R=3,4,5 then we will vote for the wrong class overall

5 wrong: $\left(\frac{1}{3}\right)^5$

4 wrong: $\underbrace{\binom{5}{4}}_{=5}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)$

3 wrong: $\underbrace{\binom{5}{3}}_{=10}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^2$

$+$

$\Rightarrow \boxed{0.21}$

$\binom{5}{4} = \frac{5!}{4! \cdot 1!}$

$= 5$

$10 = \frac{5 \cdot 4}{2}$

$\binom{5}{4} = \binom{5}{1}$

Handout 20, Q2

Handout 20, Question 2

- This analysis assumed classifiers were independent!

- What if they are not? How did Random Forests help us decorrelate classifiers?

Handout 20, Question 2

- This analysis assumed classifiers were independent!

- What if they are not? How did Random Forests help us decorrelate classifiers?

- Note about Bagging: choosing *n* with resampling actually does produce a very different dataset
  - As *n* increases, roughly 0.37 not chosen each time