



“All the hundreds of millions of people who, in their time, believed the Earth was flat never succeeded in unrounding it by an inch”

Isaac Asimov

CMPUT 365

Introduction to RL

Reminder I

You **should be enrolled in the private session** we created in Coursera for CMPUT 365.

I **cannot** use marks from the public repository for your course marks.

You **need to check, every time**, if you are in the private session and if you are submitting quizzes and assignments to the private section.

The deadlines in the public session **do not align** with the deadlines in Coursera.

If you have any questions or concerns, **talk with the TAs** or email us `cmput365@ualberta.ca`.

Reminder II / Updates

- Final Exam Schedule released
 - 12/14/2023 at 14:00 in CCIS L1-160. It will be 90 minutes long.
- The Student Perspectives of Teaching (SPOT) Survey is now available.
- The grades for the Coursera activities are available on eClass.
- Rich Sutton will give a guest lecture Dec 4th (Monday). Spread the word.



WHAT EXACTLY
IS RESEARCH?

IS GRAD SCHOOL
RIGHT FOR ME?

DEMYSTIFYING GRAD SCHOOL

HOW DO I APPLY
TO GRAD SCHOOL?

Virtual Workshop

December 6, 5-7pm

Open to all undergrad students!

Scan below to RSVP and send questions!

WHAT ARE THE
REQUIREMENTS?

ALL YOUR GRAD
SCHOOL QUESTIONS
ANSWERED HERE!



bit.ly/cs-edi-workshop23

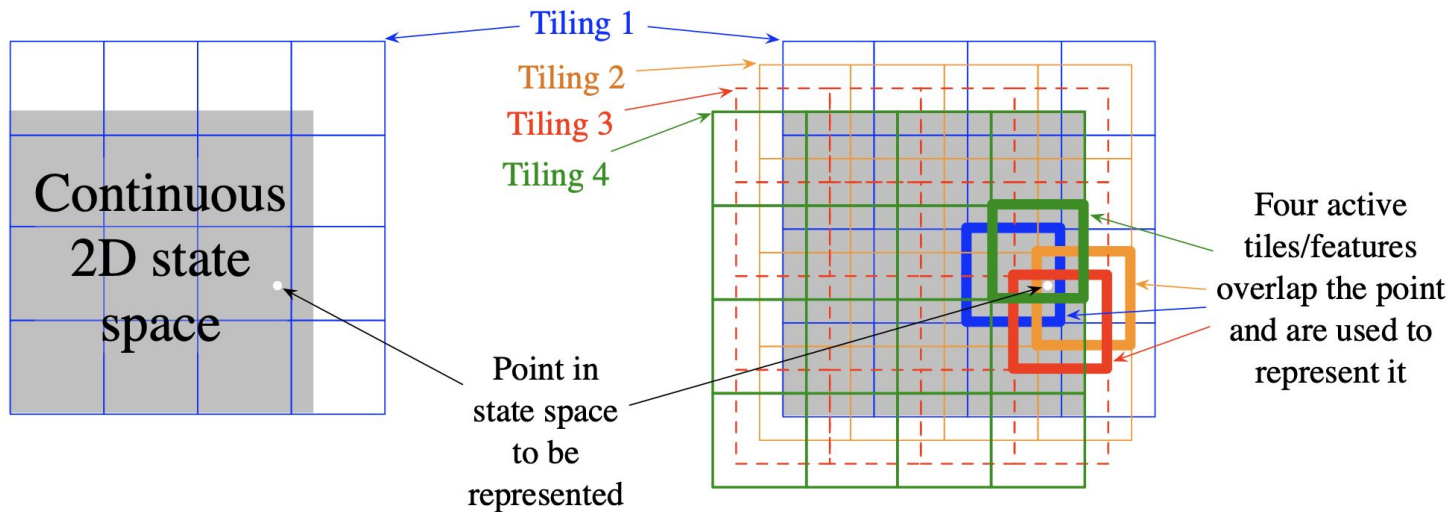


Please, interrupt me at any time!



Last Class: Tile Coding

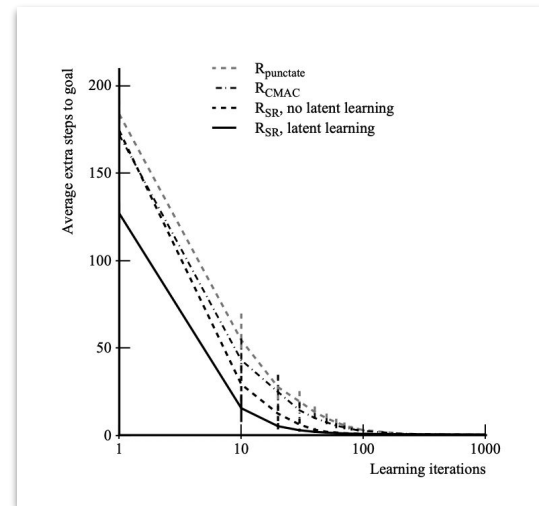
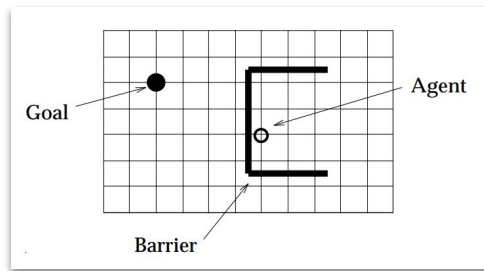
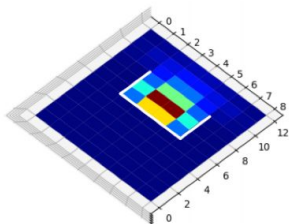
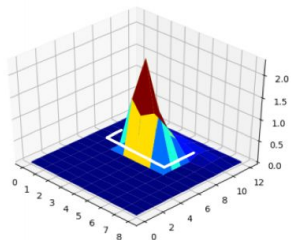
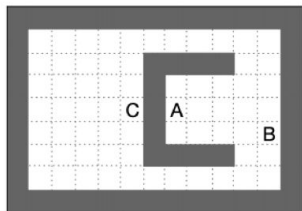
- Tile coding is a form of coarse coding for multi-dimensional continuous spaces (with a fixed number of active features per timestep).

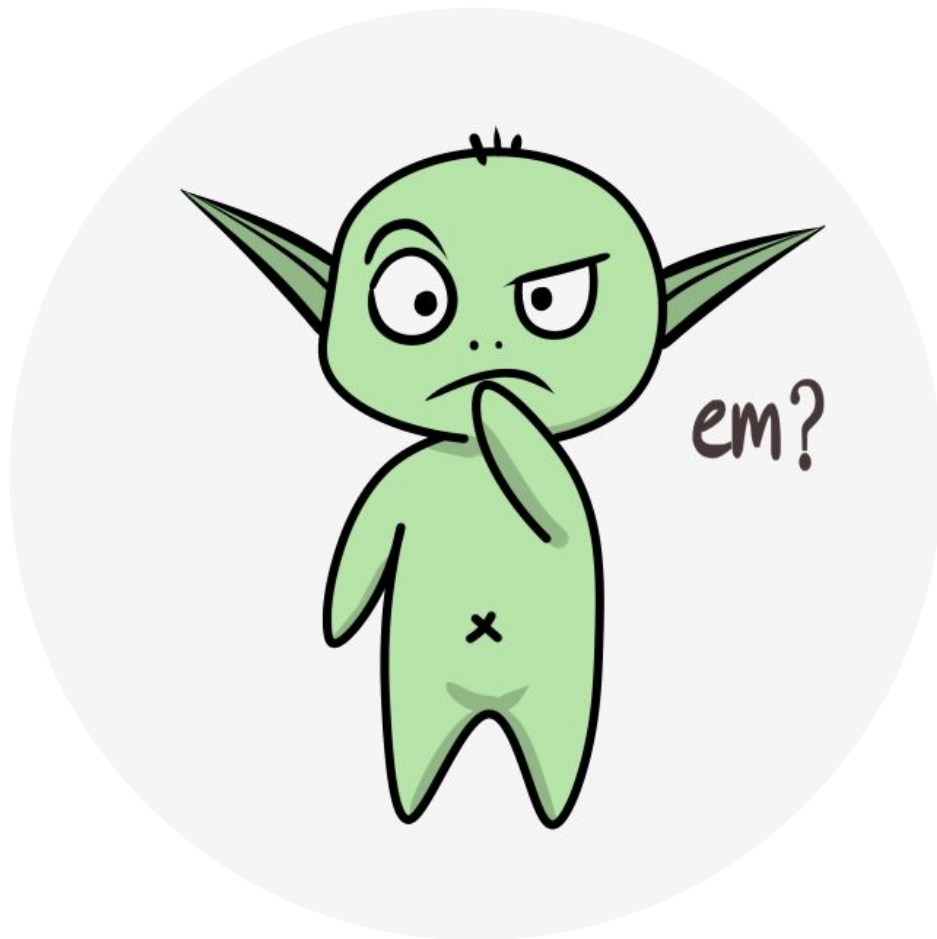


Last Class: It Isn't that We do Function Approximation Because We Cannot do Tabular Reinforcement Learning

- Successor Representation [Dayan, Neural Computation 1993].

$$\Psi_{\pi}(s, s') = \mathbb{E}_{\pi} \left[\sum_t \gamma^t \mathbf{1}_{S_t = s'} \mid S_0 = s \right]$$





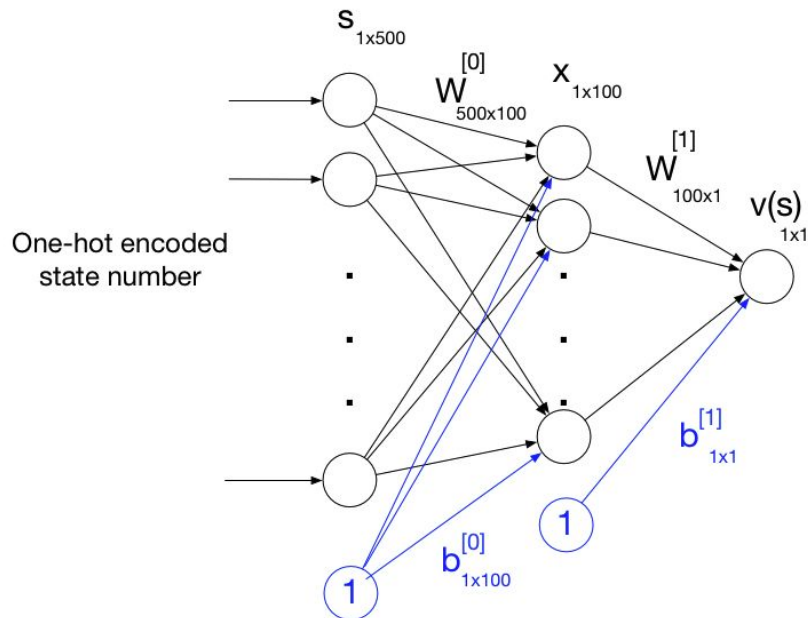
Nonlinear Function Approximation: Artificial Neural Networks

- The basics of deep reinforcement learning.
- Idea: Instead of using linear features, we feed the “raw” input to a neural network and ask it to predict the state (or state-action) value function.



Nonlinear Function Approximation: Artificial Neural Networks

The neural network looks like this:

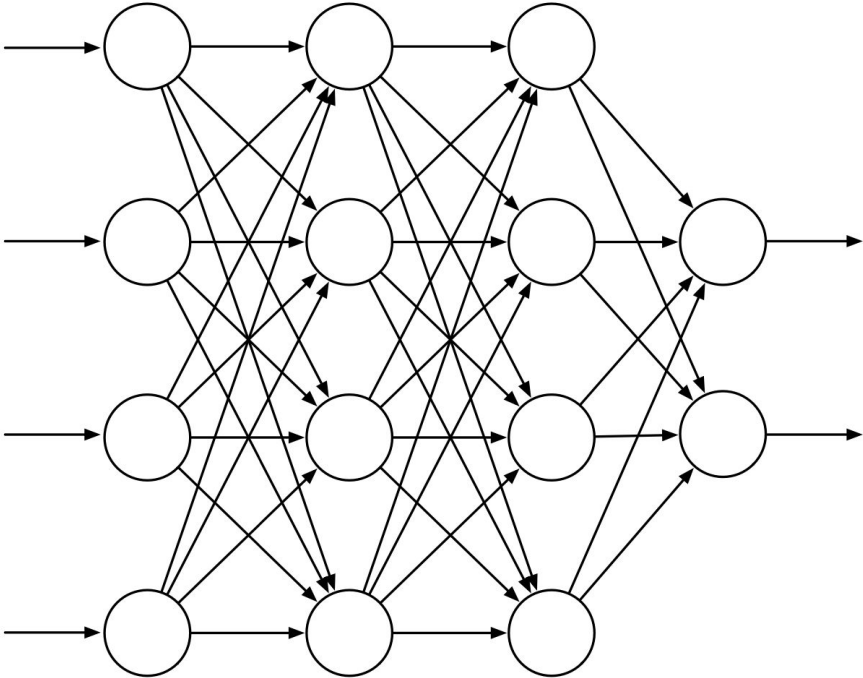


For a given input, s , value of s is computed by:

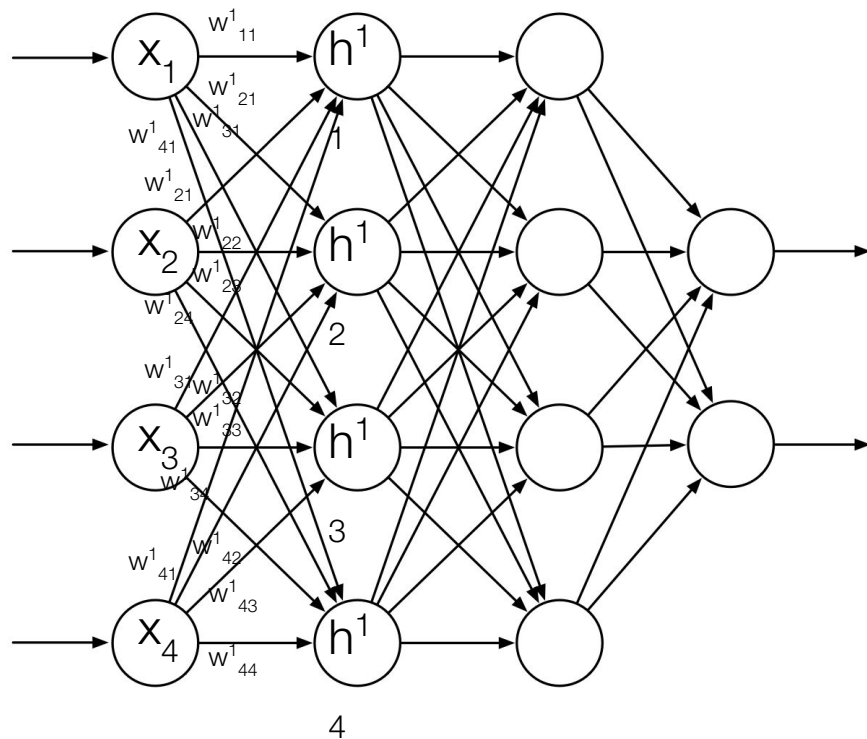
$$\begin{aligned}\psi &= sW^{[0]} + b^{[0]} \\ x &= \max(0, \psi) \\ v &= xW^{[1]} + b^{[1]}\end{aligned}$$

where $W^{[0]}$, $b^{[0]}$, $W^{[1]}$, $b^{[1]}$ are the parameters of the network and will be learned when training the agent.

Neural Networks



Neural Networks



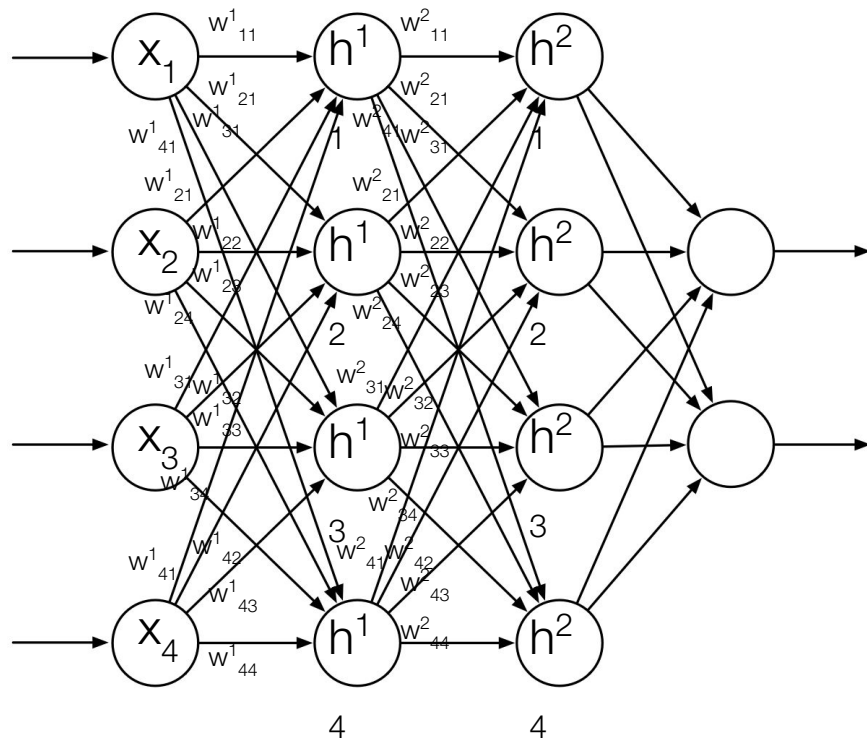
The activation function
introduces non-linearity

E.g.: $f(x) = \max(0, x)$

$$\mathbf{h}^1 = \text{act}(\mathbf{x}\mathbf{W}^1 + \mathbf{w}_1^T \mathbf{b}^1)$$

$$\text{s.t. } h^1_1 = x_1 w^1_{11} + x_2 w^1_{21} + x_3 w^1_{31} + x_4 w^1_{41} + B^1$$

Neural Networks



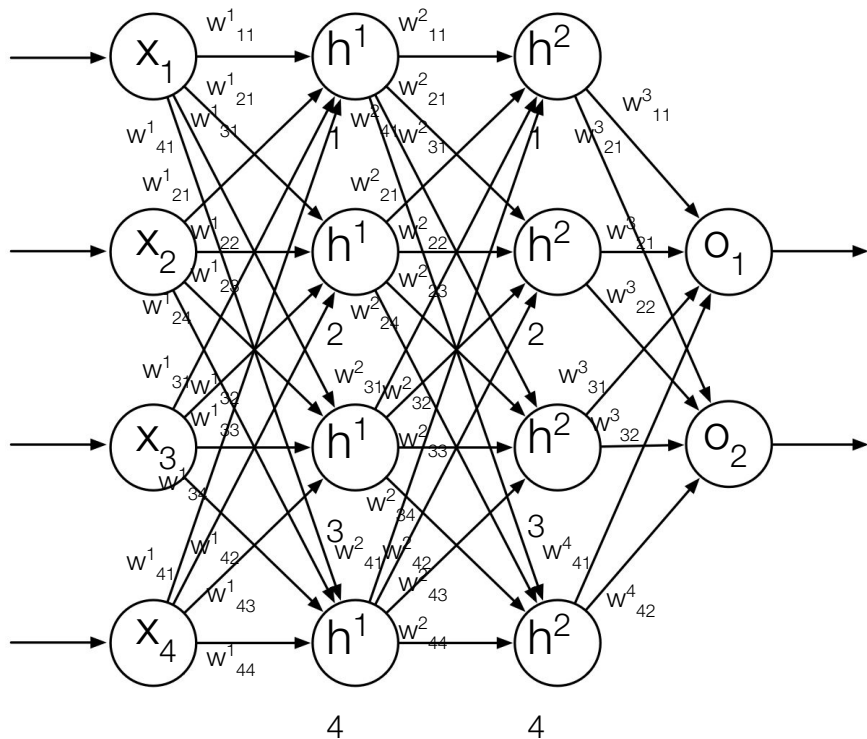
$$\mathbf{h}^1 = \text{act}(\mathbf{x}\mathbf{W}^1 + \mathbf{w}_1^T\mathbf{b}^1)$$

$$\text{s.t. } h^1_1 = x_1w^1_{11} + x_2w^1_{21} + x_3w^1_{31} + x_4w^1_{41} + B^1$$

$$\mathbf{h}^2 = \text{act}(\mathbf{h}^1\mathbf{W}^2 + \mathbf{w}_2^T\mathbf{b}^2)$$

$$\text{s.t. } h^2_1 = h^1_1w^2_{11} + h^1_2w^2_{21} + h^1_3w^2_{31} + h^1_4w^2_{41} + B^2$$

Neural Networks



$$\mathbf{h}^1 = \text{act}(\mathbf{x}\mathbf{W}^1 + \mathbf{w}_1\mathbf{b}^1)$$

$$\text{s.t. } h^1_1 = x_1w^1_{11} + x_2w^1_{21} + x_3w^1_{31} + x_4w^1_{41} + B^1$$

$$\mathbf{h}^2 = \text{act}(\mathbf{h}^1\mathbf{W}^2 + \mathbf{w}_2\mathbf{b}^2)$$

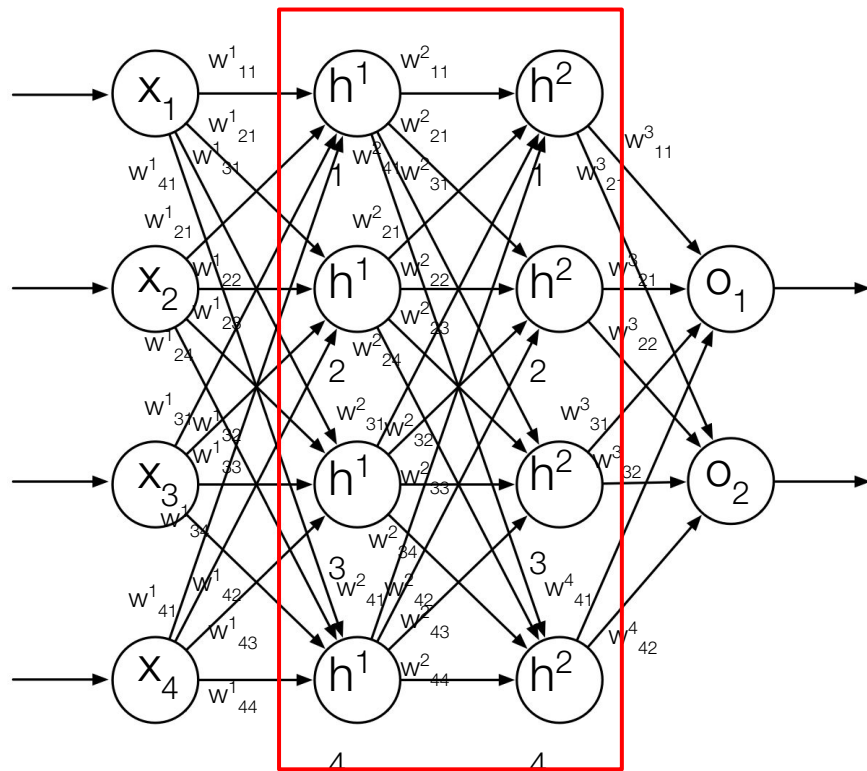
$$\text{s.t. } h^2_1 = h^1_1w^2_{11} + h^1_2w^2_{21} + h^1_3w^2_{31} + h^1_4w^2_{41} + B^2$$

$$\mathbf{o} = \text{act}(\mathbf{h}^2\mathbf{W}^3 + \mathbf{w}_3\mathbf{b}^3)$$

$$\text{s.t. } o_1 = h^2_1w^3_{11} + h^2_2w^3_{21} + h^2_3w^3_{31} + h^2_4w^3_{41} + B^3$$

$$\mathbf{o} = \text{act}(\text{act}(\text{act}(\mathbf{x}\mathbf{W}^1 + \mathbf{w}_1\mathbf{b}^1)\mathbf{W}^2 + \mathbf{w}_2\mathbf{b}^2)\mathbf{W}^3 + \mathbf{w}_3\mathbf{b}^3)$$

Neural Networks



**Representation
(Learned features)**

$$\mathbf{h}^1 = \text{act}(\mathbf{x}\mathbf{W}^1 + \mathbf{w}_1\mathbf{T}\mathbf{b}^1)$$

$$\text{s.t. } h^1_1 = x_1w^1_{11} + x_2w^1_{21} + x_3w^1_{31} + x_4w^1_{41} + B^1$$

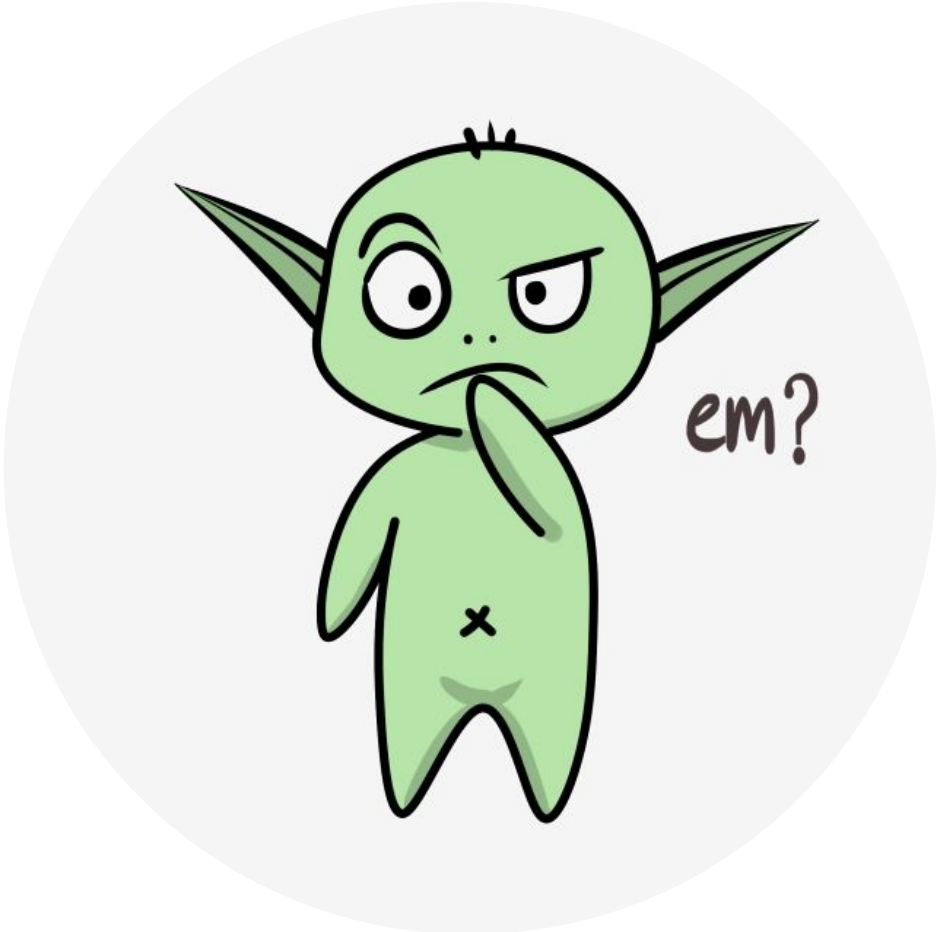
$$\mathbf{h}^2 = \text{act}(\mathbf{h}^1\mathbf{W}^2 + \mathbf{w}_2\mathbf{T}\mathbf{b}^2)$$

$$\text{s.t. } h^2_1 = h^1_1w^2_{11} + h^1_2w^2_{21} + h^1_3w^2_{31} + h^1_4w^2_{41} + B^2$$

$$\mathbf{o} = \text{act}(\mathbf{h}^2\mathbf{W}^3 + \mathbf{w}_3\mathbf{T}\mathbf{b}^3)$$

$$\text{s.t. } o_1 = h^2_1w^3_{11} + h^2_2w^3_{21} + h^2_3w^3_{31} + h^2_4w^3_{41} + B^3$$

$$\mathbf{o} = \text{act}(\text{act}(\text{act}(\mathbf{x}\mathbf{W}^1 + \mathbf{w}_1\mathbf{T}\mathbf{b}^1)\mathbf{W}^2 + \mathbf{w}_2\mathbf{T}\mathbf{b}^2)\mathbf{W}^3 + \mathbf{w}_3\mathbf{T}\mathbf{b}^3)$$



A Note from the Textbook

The backpropagation algorithm can produce good results for shallow networks having 1 or 2 hidden layers, but it may not work well for deeper ANNs. In fact, training a network with $k + 1$ hidden layers can actually result in poorer performance than training a network with k hidden layers, even though the deeper network can represent all the functions that the shallower network can (Bengio, 2009). Explaining results like these is not easy, but several factors are important. First, the large number of weights in a typical deep ANN makes it difficult to avoid the problem of overfitting, that is, the problem of failing to generalize correctly to cases on which the network has not been trained. Second, backpropagation does not work well for deep ANNs because the partial derivatives computed by its backward passes either decay rapidly toward the input side of the network, making learning by deep layers extremely slow, or the partial derivatives grow rapidly toward the input side of the network, making learning unstable. Methods

Things change quickly...

Deep Convolutional Network

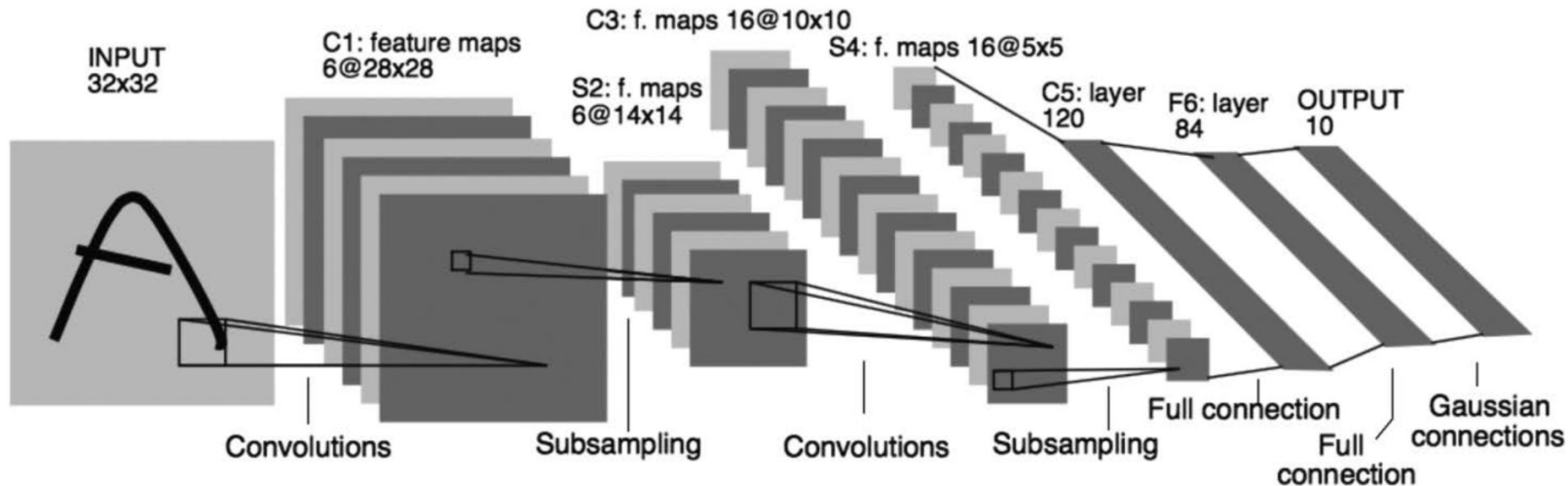
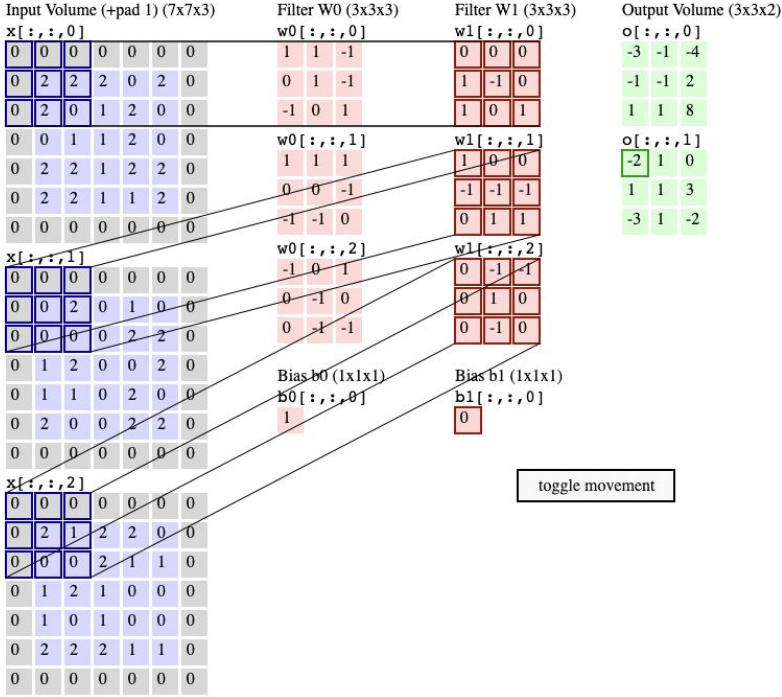
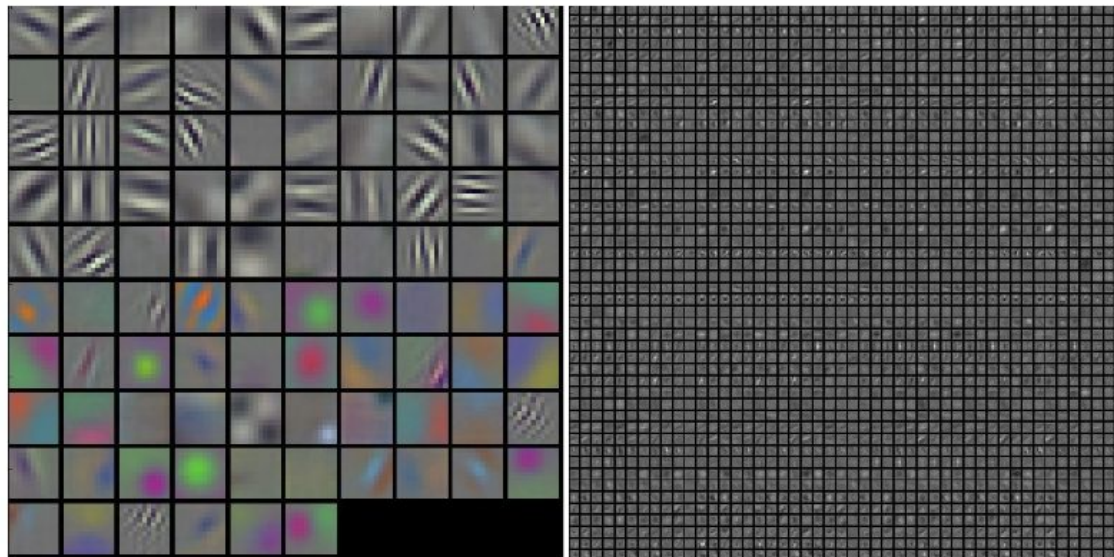


Figure 9.15: Deep Convolutional Network. Republished with permission of Proceedings of the IEEE, from Gradient-based learning applied to document recognition, LeCun, Bottou, Bengio, and Haffner, volume 86, 1998; permission conveyed through Copyright Clearance Center, Inc.

Deep Convolutional Network



Learned Representations



Typical-looking filters on the first CONV layer (left), and the 2nd CONV layer (right) of a trained AlexNet. Notice that the first-layer weights are very nice and smooth, indicating nicely converged network. The color/grayscale features are clustered because the AlexNet contains two separate streams of processing, and an apparent consequence of this architecture is that one stream develops high-frequency grayscale features and the other low-frequency color features. The 2nd CONV layer weights are not as interpretable, but it is apparent that they are still smooth, well-formed, and absent of noisy patterns.

