CS540 Introduction to Artificial Intelligence
Lecture 8

Young Wu
Based on lecture slides by Jerry Zhu, Yingyu Liang, and Charles Dyer

June 8, 2020

SIFT and HOG Features

Motivation

- SIFT and HOG features are expensive to compute.
- Simpler features should be used for real time face detection tasks.

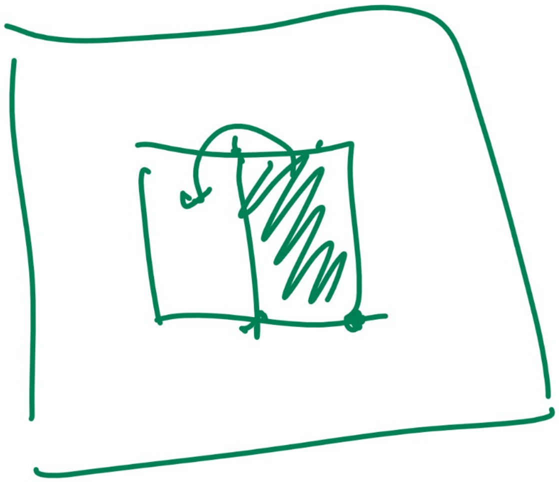
Real Time Face Detection

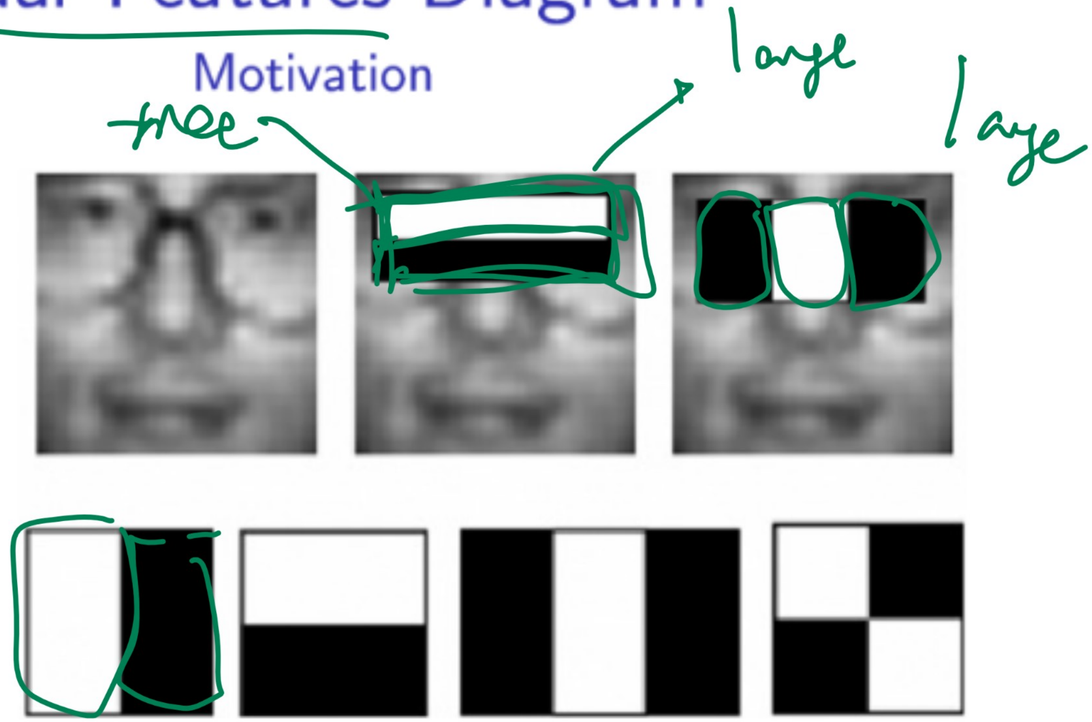
Motivation

- Each image contains 10000 to 500000 locations and scales.
- Faces occur in 0 to 50 per image.
- Want a very small number of false positives.

Haar Features Diagram

Motivation





Weak Classifiers

Definition

- Each weak classifier is a decision stump (decision tree with only one split) using one Haar feature x.

$$\boxed{f(x) = \mathbb{1}_{\{x>\theta\}}}$$

- Finding the threshold by comparing the information gain from all possible splits is too expensive, so θ is usually computed as the average of the mean values of the feature for each class.

non face   face

$$\theta = \frac{1}{2}\left(\frac{1}{n_0}\sum_{i:y_i=0} x_i + \frac{1}{n_1}\sum_{i:y_i=1} x_i\right)$$

Strong Classifiers

Definition

- The weak classifiers are trained sequentially using ensemble methods such as AdaBoost.
- A sequence of T weak classifiers is called aT -strong classifier.
- Multiple T -strong classifiers can be trained for different values of T and combined into a cascaded classifier.

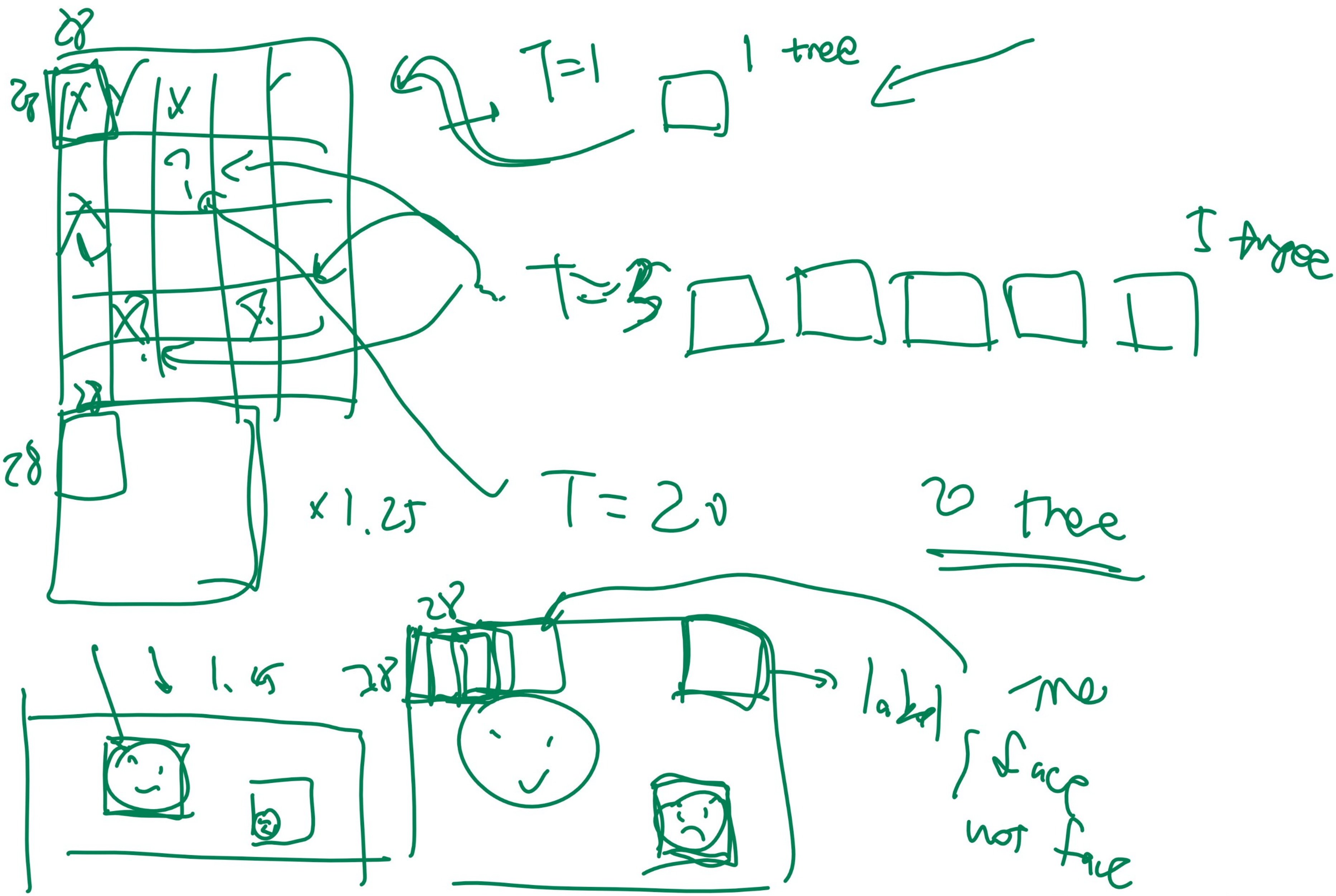
Cascaded Classifiers

Definition

- Start with aT -strong classifier with small T, and use it reject obviously negative regions (regions with no faces).
- Train and use aT -strong classifier with larger T on only the regions that are not rejected.
- Repeat this process with stronger classifiers.

Viola Jones Diagram

Discussion



Convolution Example

Quiz

middle element

- Find the gradient magnitude and direction for the center cell of the following image. Use the derivative filters $\begin{bmatrix}-1\\0\\1\end{bmatrix}$ and $\begin{bmatrix}-1 & 0 & 1\end{bmatrix}$
- $\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix}$

zero-padding

$$\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} \cdot \begin{bmatrix}1\\0\\-1\end{bmatrix} = \begin{bmatrix}-4 & -5 & -6\\-6 & \boxed{-6} & -6\\4 & 5 & 6\end{bmatrix} \to D_y I$$

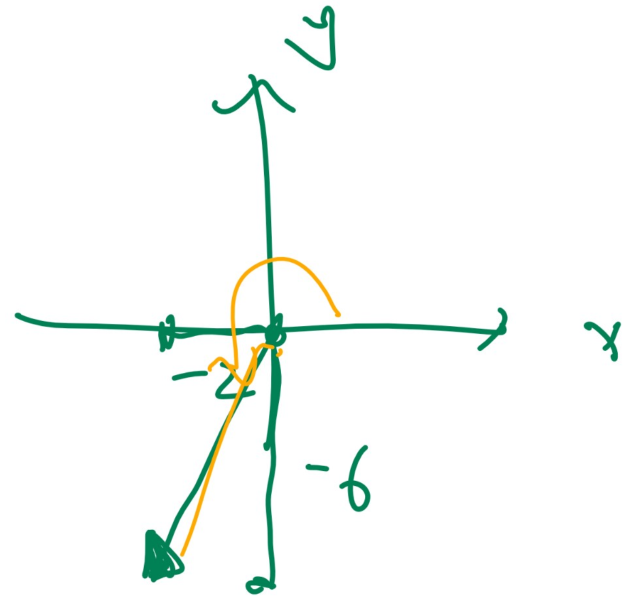
$$\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} \cdot \begin{bmatrix}1 & 0 & -1\end{bmatrix} = \begin{bmatrix}-2 & -2 & 2\\-5 & \boxed{-2} & 5\\-8 & -2 & 8\end{bmatrix} \to D_x I$$

Gradient Example

Quiz

$$G = \sqrt{(-6)^2 + (-2)^2} = \ldots$$

gradient magnitude



$$\theta = \arctan\left(\frac{6}{2}\right)$$

Convolution Example 1

Quiz

left Sobel

$$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} * \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

$D_x I$

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} = F^{flip}$$

3×3

- A: $\begin{bmatrix} -1 & -3 & -3 \\ 0 & 0 & 0 \\ 1 & 3 & 3 \end{bmatrix}$, B : $\begin{bmatrix} -3 & -3 & 3 \\ -4 & -4 & 4 \\ -3 & -3 & 3 \end{bmatrix}$ ✓
- C: $\begin{bmatrix} -3 & \cancel{-4} & -3 \\ -3 & -4 & -3 \\ 3 & 4 & 3 \end{bmatrix}$, D : $\begin{bmatrix} -1 & 0 & 1 \\ -3 & 0 & 3 \\ -3 & 0 & 3 \end{bmatrix}$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \xrightarrow{\text{flip}} \begin{bmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix}$$

Convolution Example 2

Quiz

Q2

back 6:45

on midterm

bottom Sobel

$$\begin{bmatrix}0&1&1\\0&1&1\\0&1&1\end{bmatrix} * \begin{bmatrix}-1&-2&-1\\0&0&0\\1&2&1\end{bmatrix}$$

$$* \begin{bmatrix}-1\\0\\1\end{bmatrix} * \begin{bmatrix}1&2&1\end{bmatrix}$$

implement a lot faster

- A: $\begin{bmatrix}-1&-3&-3\\0&0&0\\1&3&3\end{bmatrix}$, B : $\begin{bmatrix}-3&-3&3\\-4&-4&4\\-3&-3&3\end{bmatrix}$
- C: $\begin{bmatrix}-3&-4&-3\\-3&-4&-3\\3&4&3\end{bmatrix}$, D : $\begin{bmatrix}-1&0&1\\-3&0&3\\-3&0&3\end{bmatrix}$

flip $\begin{bmatrix}1&2&1\\0&0&0\\-1&-2&-1\end{bmatrix}$

$$\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix} \xrightarrow{\text{flip}} \begin{bmatrix}9&8&7\\6&5&4\\3&2&1\end{bmatrix}$$

Convolution Example 3

Quiz

Q3

$$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} * \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} * \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}$$

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

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

- A: $\begin{bmatrix} -1 & -3 & -3 \\ 0 & 0 & 0 \\ 1 & 3 & 3 \end{bmatrix}$, B : $\begin{bmatrix} -3 & -3 & 3 \\ -4 & -4 & 4 \\ -3 & -3 & 3 \end{bmatrix}$
- C: $\begin{bmatrix} -3 & -4 & -3 \\ -3 & -4 & -3 \\ 3 & 4 & 3 \end{bmatrix}$, D : $\begin{bmatrix} -1 & 0 & 1 \\ -3 & 0 & 3 \\ -3 & 0 & 3 \end{bmatrix}$

Convolution Example 4

Quiz

Q4

What is the gradient magnitude for the center ~~cell~~ element?

$$\nabla_x = \begin{bmatrix} -3 & -3 & 3 \\ -4 & -4 & 4 \\ -3 & -3 & 3 \end{bmatrix}, \nabla_y = \begin{bmatrix} -1 & -3 & -3 \\ 0 & 0 & 0 \\ 1 & 3 & 3 \end{bmatrix}$$

- A: 1, B: 2, C: 3, D: 4, E: 5

$$\sqrt{-4^2 + 0^2} = 4$$

Convolution Example 5

Quiz

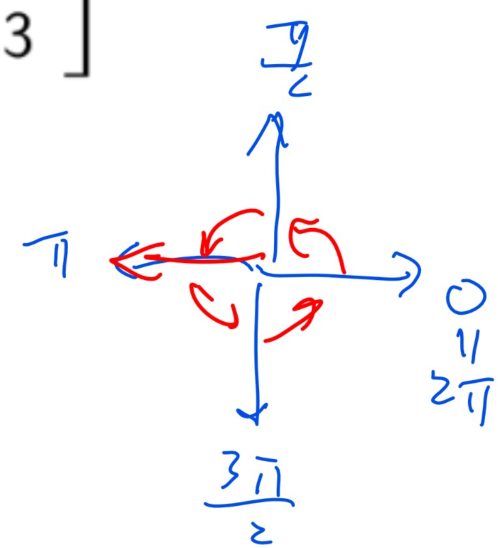
Q5

not on midterm

What is the gradient direction bin for the center ~~cell~~ element?

$$\nabla_x = \begin{bmatrix} -3 & -3 & 3 \\ -4 & -4 & 4 \\ -3 & -3 & 3 \end{bmatrix}, \nabla_y = \begin{bmatrix} -1 & -3 & -3 \\ 0 & 0 & 0 \\ 1 & 3 & 3 \end{bmatrix}$$

- A: $\left(0, \frac{\pi}{2}\right]$, $\boxed{\text{B:}}$ $\left(\frac{\pi}{2}, \pi\right]$, C: $\left(\pi, 3\frac{\pi}{2}\right]$, D: $\left(3\frac{\pi}{2}, 2\pi\right]$

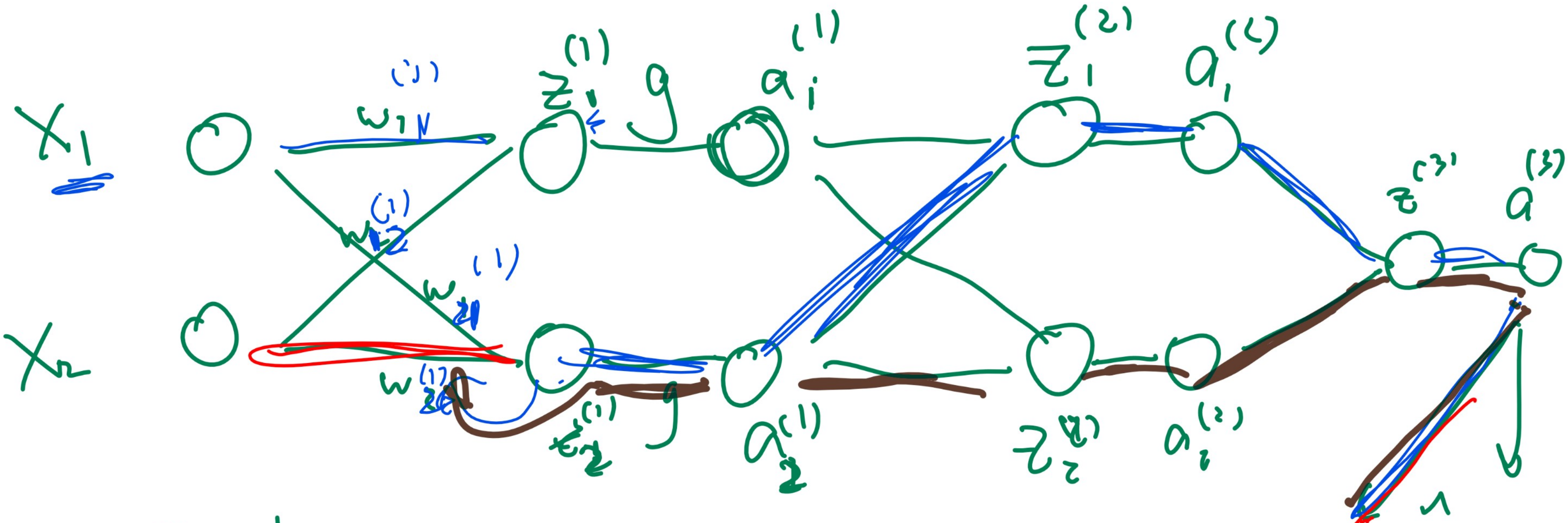


atan2(dy, dx)

−4

Three Layer Neural Network Weights Diagram 1

Quiz



$z = w^T x + b \qquad a = g(z)$

Cost $\hat{y}_i$

$w = w - \alpha \frac{\partial C}{\partial w}$

$$\frac{\partial C}{\partial w^{(1)}_{22}} = \sum_{i=1}^{n} \frac{\partial C}{\partial a^{(3)}} \cdot \frac{\partial a^{(3)}}{\partial z^{(3)}} \cdot \frac{\partial z^{(3)}}{\partial a^{(2)}_1} \cdot \frac{\partial a^{(2)}_1}{\partial z^{(2)}_1} \cdot \frac{\partial z^{(2)}_1}{\partial a^{(1)}_2} \cdot \frac{\partial a^{(1)}_2}{\partial z^{(1)}_2} \cdot \frac{\partial z^{(1)}_2}{\partial w^{(1)}_{22}}$$

Three Layer Neural Network Weights Diagram 2

Quiz

$$+ \frac{\partial C}{\partial a^{(3)}} \cdot \frac{\partial a^{(3)}}{\partial z^{(3)}} \cdot \frac{\partial z^{(3)}}{\partial a_2^{(2)}} \cdot \frac{\partial a_2^{(2)}}{\partial z_2^{(2)}} \cdot \frac{\partial z_2^{(2)}}{\partial a_2^{(1)}} \cdot \frac{\partial a_2^{(1)}}{\partial z_2^{(1)}} \cdot \frac{\partial z_2^{(1)}}{\partial w_{22}^{(1)}}$$

$$C = \frac{1}{2}(y - a)^2$$

$$(a^{(3)} - y_i)\,(a^{(3)})(1 - a^{(3)}) \cdot w_{21}^{(2)} \quad a_2^{(2)}(1 - a_2^{(2)})$$

$$g' = g \cdot (1 - g) \text{ for sigmoid}$$

$$w_{22}^{(1)} \cdot a_2^{(1)}(1 - a_2^{(1)})\, x_2$$

Three Layer Neural Network Backpropogation

Quiz

- Which of the following is correct for a three layer network? Assume there are 10 units in the first layer and 5 units in the second layer.
- Choices on the next page.

Three Layer Neural Network Backpropogation

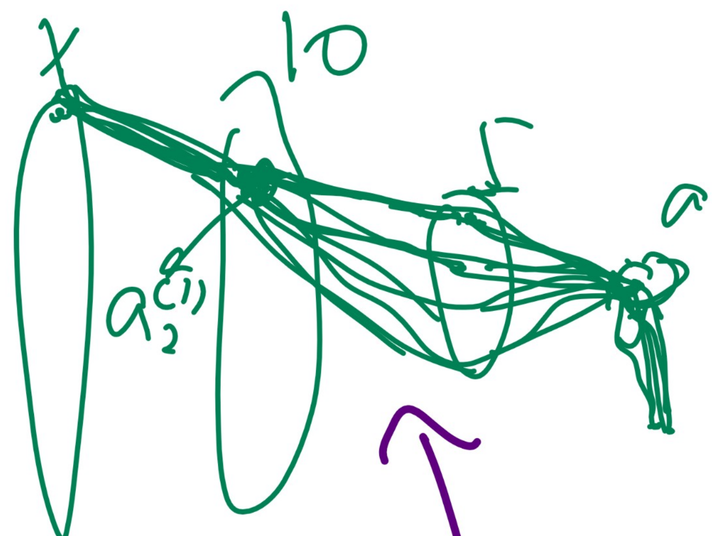
Quiz

Q6

- ~~A~~: $$\frac{\partial C}{\partial w_{12}^{(1)}} = \sum_{j'=1}^{10}\sum_{j=1}^{5} \frac{\partial C}{\partial a_j^{(2)}} \frac{\partial a_j^{(2)}}{\partial a_{j'}^{(1)}} \frac{\partial a_{j'}^{(1)}}{\partial w_{1j'}^{(1)}}$$
- B: $$\frac{\partial C}{\partial w_{12}^{(1)}} = \sum_{j=1}^{5} \frac{\partial C}{\partial a_j^{(2)}} \frac{\partial a_j^{(2)}}{\partial a_1^{(1)}} \frac{\partial a_1^{(1)}}{\partial w_{12}^{(1)}}$$
- C: $$\frac{\partial C}{\partial w_{12}^{(1)}} = \sum_{j=1}^{5} \frac{\partial C}{\partial a_j^{(2)}} \frac{\partial a_j^{(2)}}{\partial a_2^{(1)}} \frac{\partial a_2^{(1)}}{\partial w_{12}^{(1)}}$$
- ~~D~~: $$\frac{\partial C}{\partial w_{12}^{(1)}} = \frac{\partial C}{\partial a_1^{(2)}} \frac{\partial a_1^{(2)}}{\partial a_1^{(1)}} \frac{\partial a_1^{(1)}}{\partial w_{12}^{(1)}}$$
- ~~E~~: $$\frac{\partial C}{\partial w_{12}^{(1)}} = \frac{\partial C}{\partial a_2^{(2)}} \frac{\partial a_2^{(2)}}{\partial a_2^{(1)}} \frac{\partial a_2^{(1)}}{\partial w_{12}^{(1)}}$$

10 hidden 1

5 hidden 2

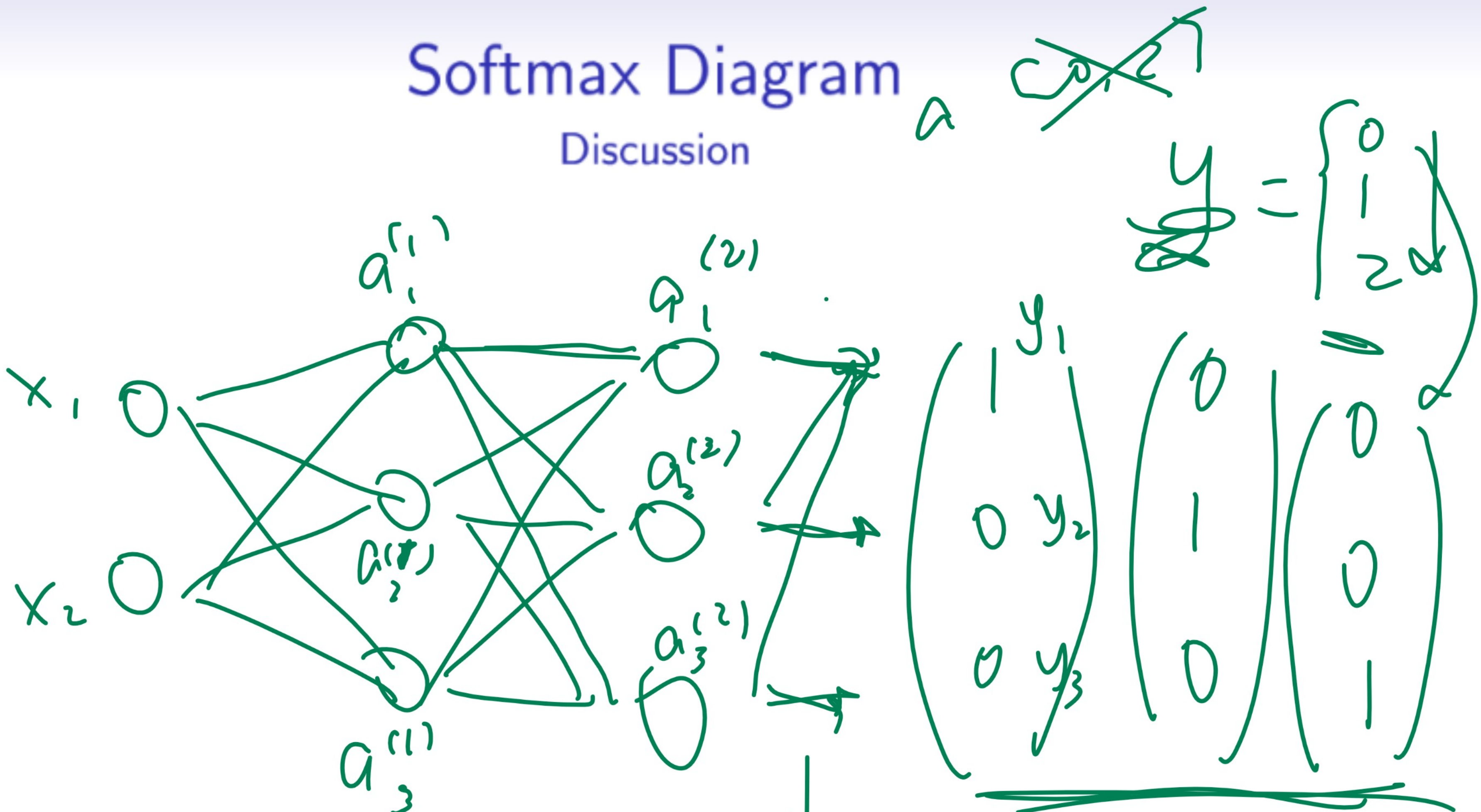


$a > 0.5 \rightarrow 1$

$< 0.5 \rightarrow 0$

Softmax Diagram

Discussion



multiclass logistic

softmax

one-hot encoding

Weight Count

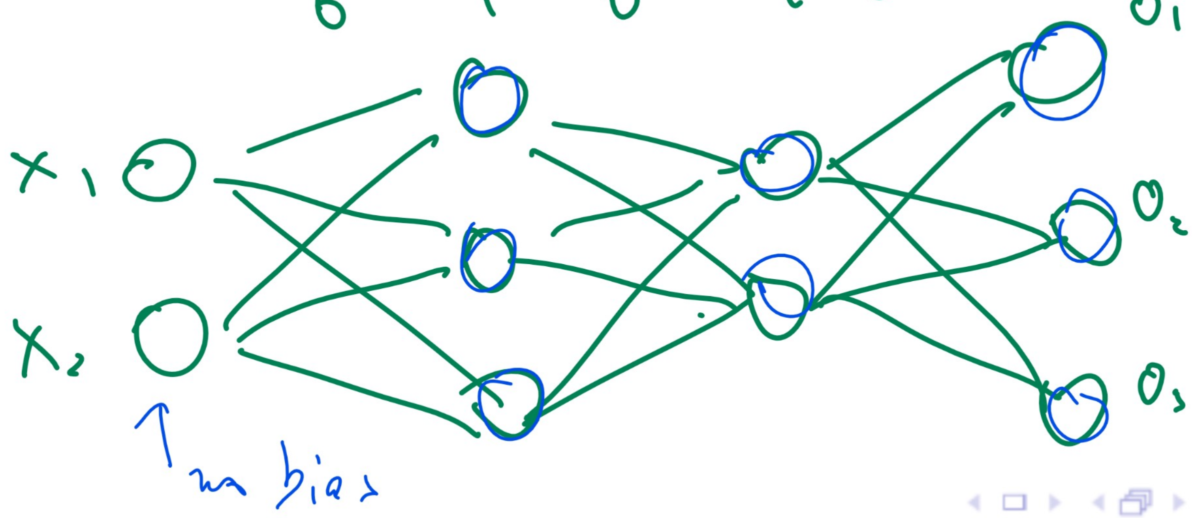
Quiz

- How many weights and biases are there in a (fully connected) three layer neural network with 2 input units, 3 hidden units in the first hidden layer, 2 hidden units in the second hidden layer, and 3 output units?

$6 + 6 + 6 = 18$ weights

8 biases

26 things to update

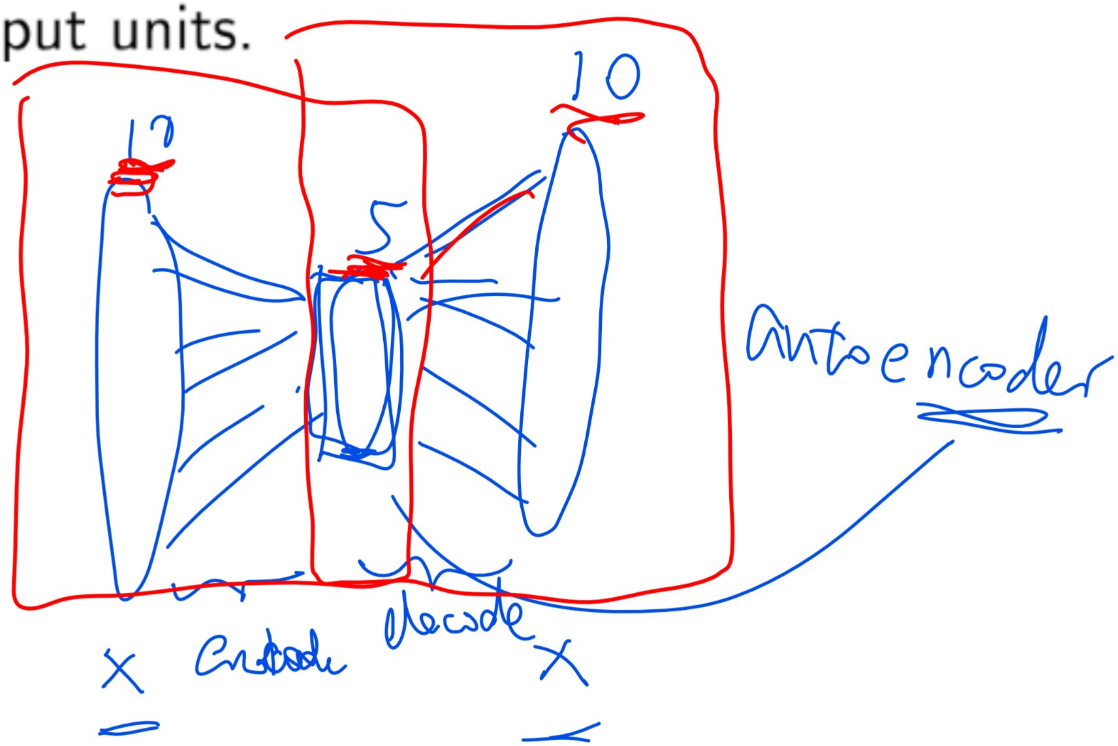


Weight Count 2

Quiz

Q7

- How many weights (not including bias) are there in a (fully connected) two layer neural network with 10 input units, 5 hidden units, and 10 output units.
- A: 50
- B: 55
- C: 100 $= 5 \cdot 10 + 10 \cdot 5$
- D: 110
- E: 500



Weight Count 3

Quiz

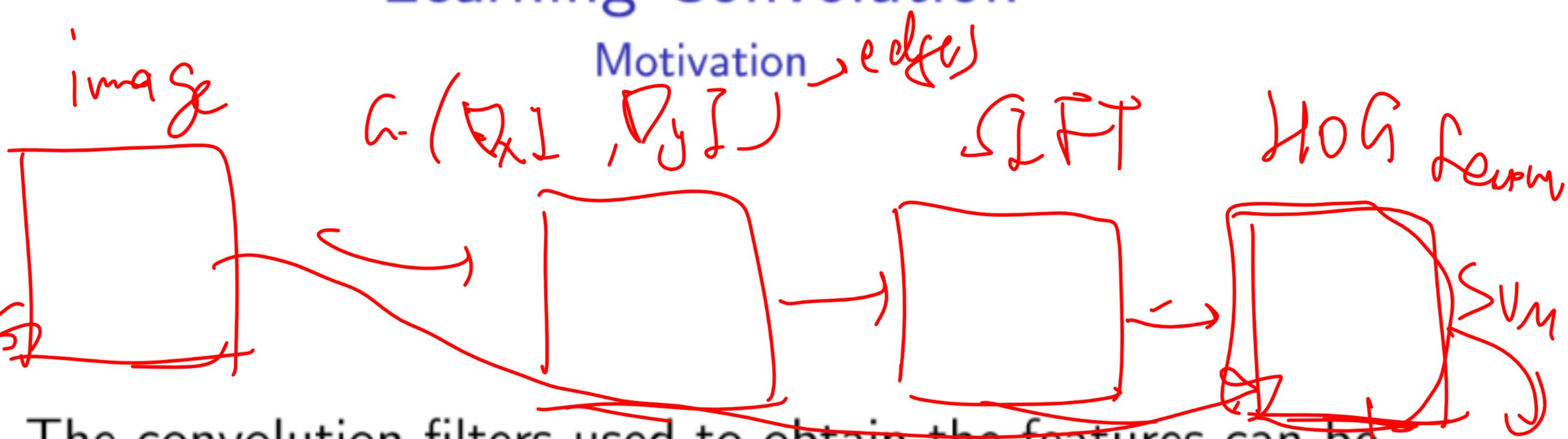
Q8

- How many biases are there in a (fully connected) two layer neural network with 10 input units, 5 hidden units, and 10 output units.
- A: 5
- B: 10
- C: 15
- D: 20
- E: 25

5 + 10 = 15

Learning Convolution

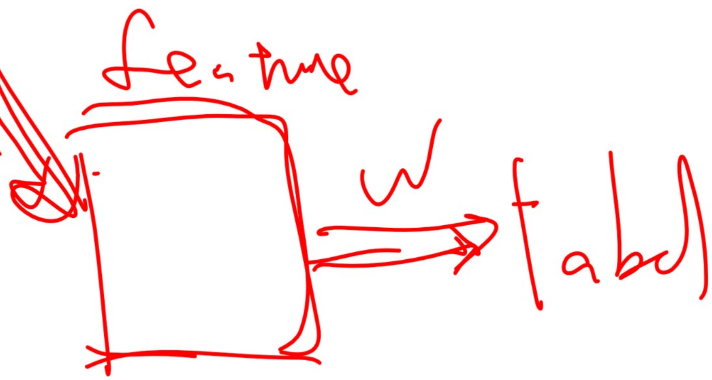
Motivation



- The convolution filters used to obtain the features can be learned in a neural network. Such networks are called convolutional neural networks and they usually contain multiple convolutional layers with fully connected and softmax layers near the end.

W
convolution

feature engineering



Convolutional Layers

Definition

- In the (fully connected) neural networks discussed previously, each input unit is associated with a different weight.

$$a = g\left(w^T x + b\right)$$

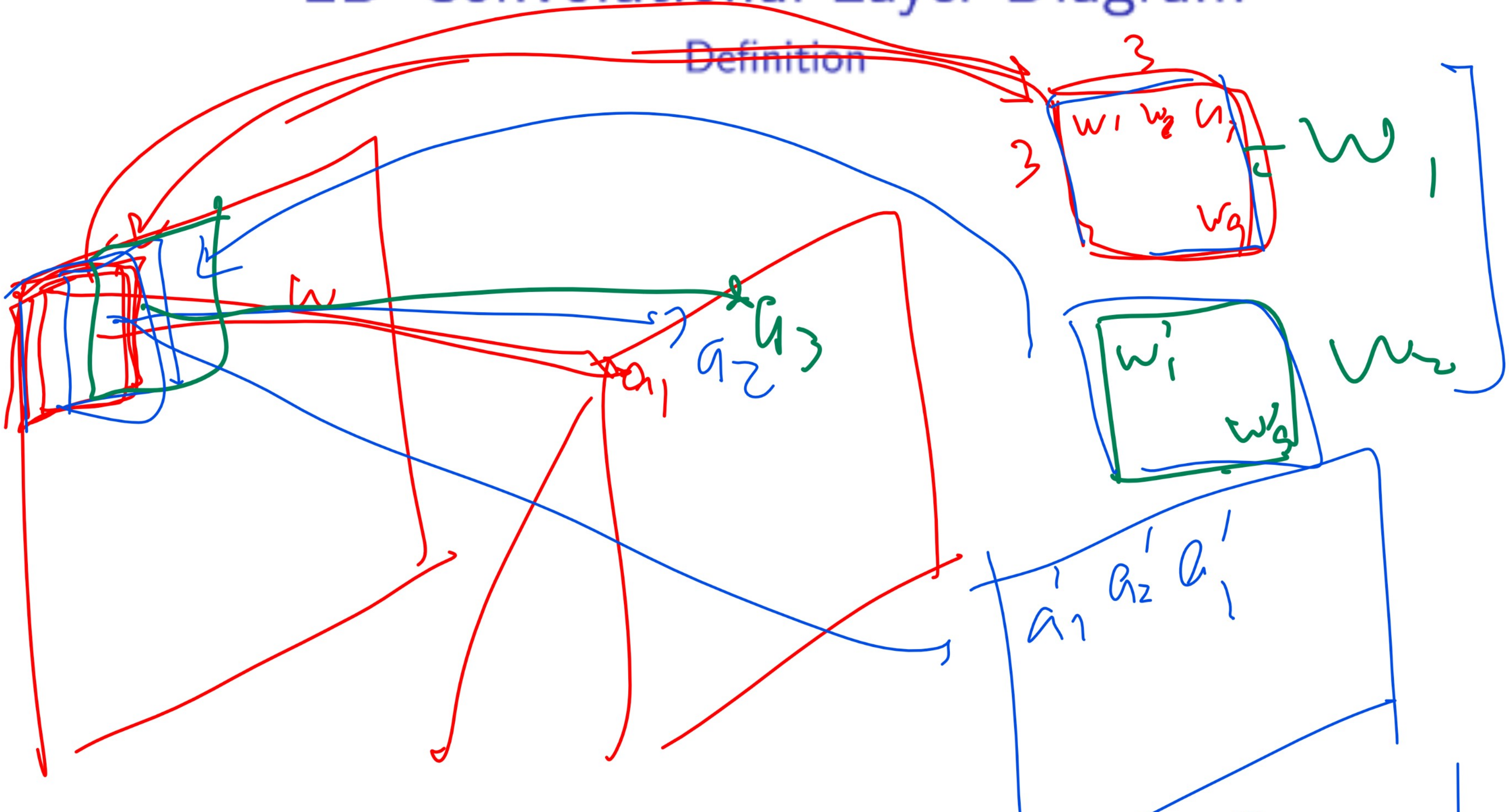
- In the convolutional layers, one single filter (a multi-dimensional array of weights) is used for all units (arranged in an array the same size as the filter).

$$A = g(W * X + b)$$

filter

2D Convolutional Layer Diagram

Definition

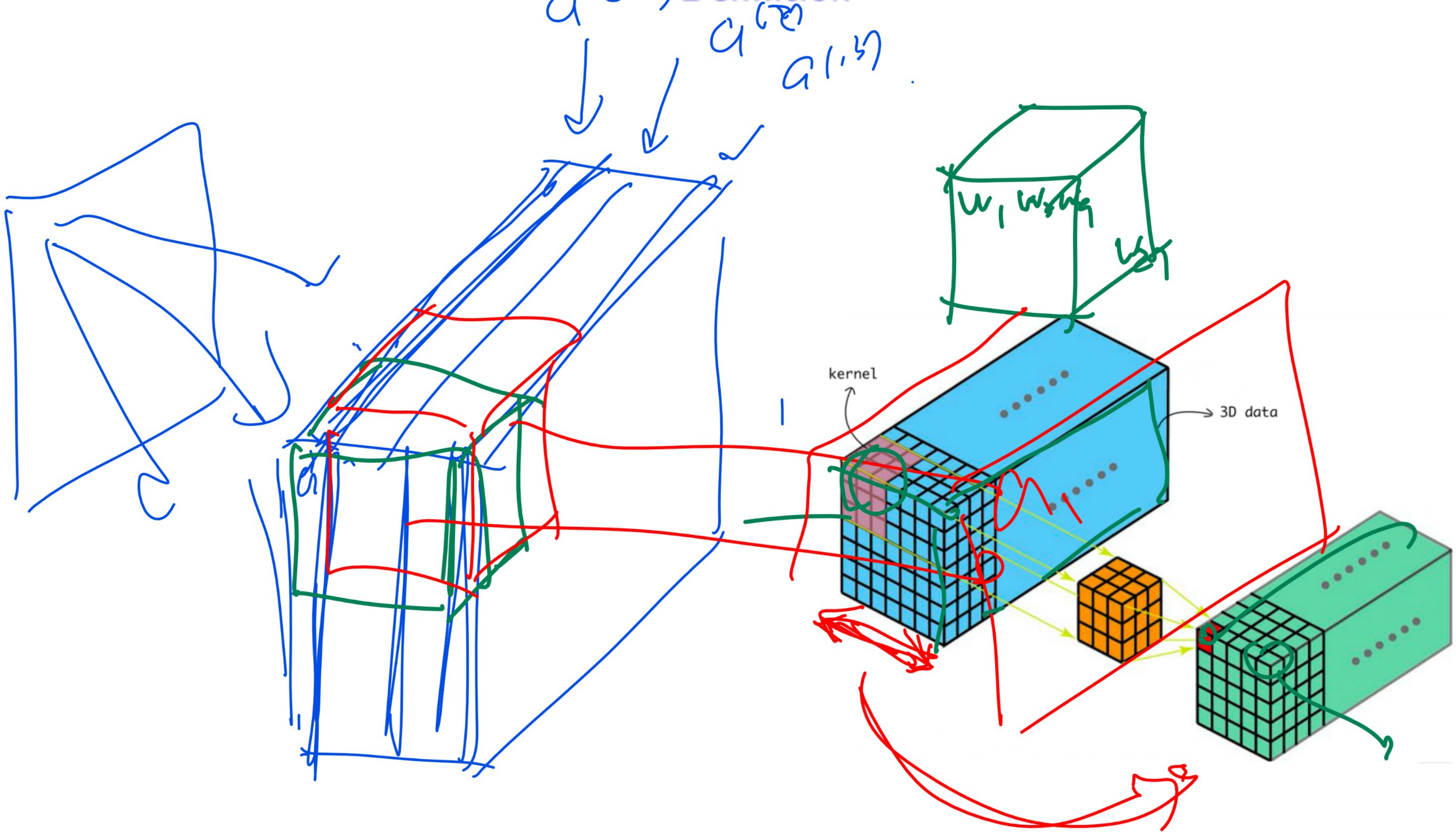


Same layer

$g(W * X(3\times 3) + b)$

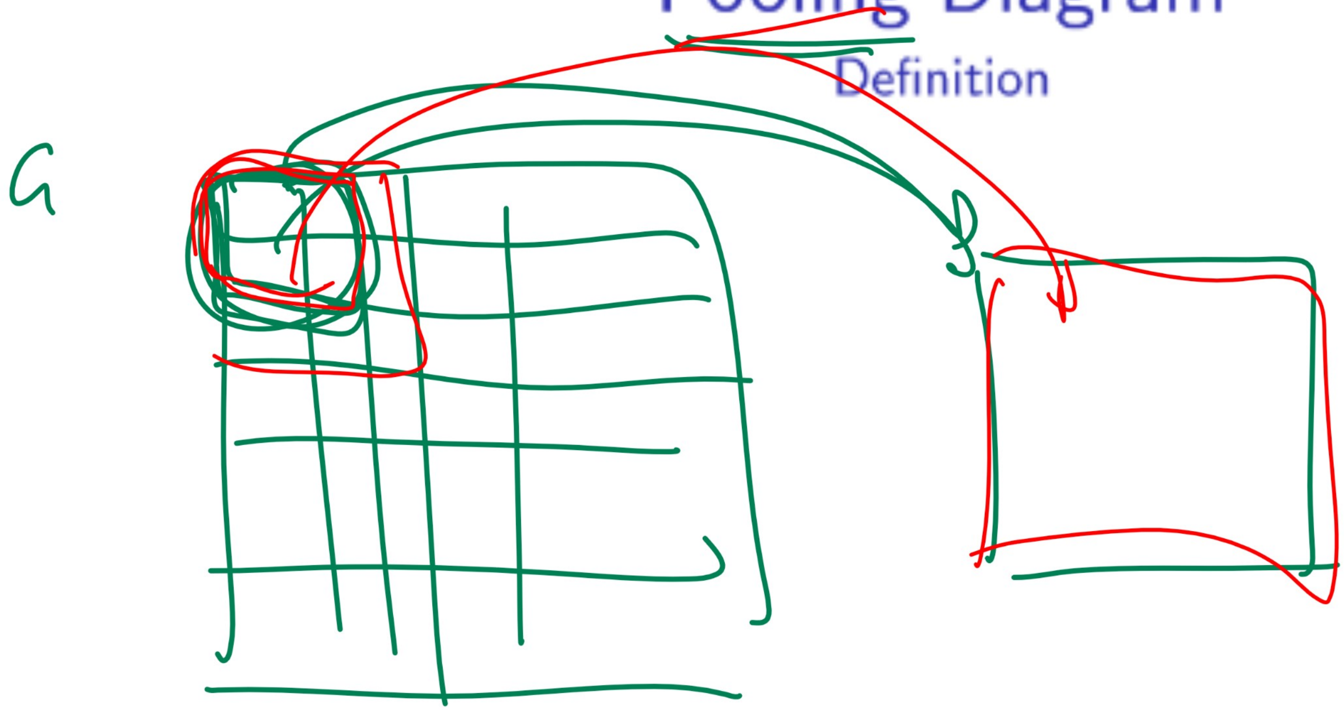
3D Convolutional Layer Diagram

Definition



Pooling Diagram

Definition



max pooling

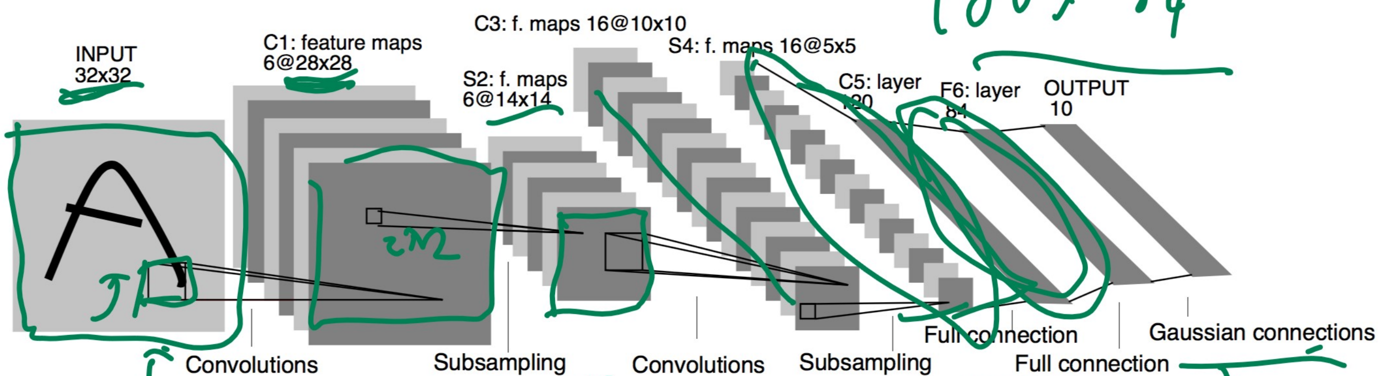
average pooling

LeNet Diagram and Demo

12

Discussion

120 x 84



softmax

25 0 25 x 6

25 x 3

$$z = W * X + b$$

$$g(z)_{ij} = g(z_{ij})$$ matrix