CS540 Introduction to Artificial Intelligence
Lecture 7

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

June 30, 2021

Hat Game

Quiz

R1

- 5 kids are wearing either green or red hats in a party: they can see every other kid's hat but not their own.
- Dad said to everyone: at least one of you is wearing a green hat.
- Dad asked everyone: do you know the color of your hat?
- Everyone said no. ← at least 2
- Dad asked again: do you know the color of your hat?
- Everyone said no. ← at least 3
- Dad asked again: do you know the color of your hat?
- Some kids (at least one) said yes. ←
- No one lied. How many kids are wearing green hats?
- A: 1... B: 2... C: 3... D: 4... E: 5

Remind Me to Start Recording

Admin

- The messages you send in chat will be recorded: you can change your Zoom name now before I start recording.

Decision Tree

Description

- Find the feature that is the most informative.
- Split the training set into subsets according to this feature.
- Repeat on the subsets until all the labels in the subset are the same.

ID3 Algorithm (Iterative Dichotomiser 3)

- The most informative feature X_j has the largest information gain:

$$I\left(Y|X_j\right) = H\left(Y\right) - H\left(Y|X_j\right)$$

example on Friday

Pruning Diagram

Discussion

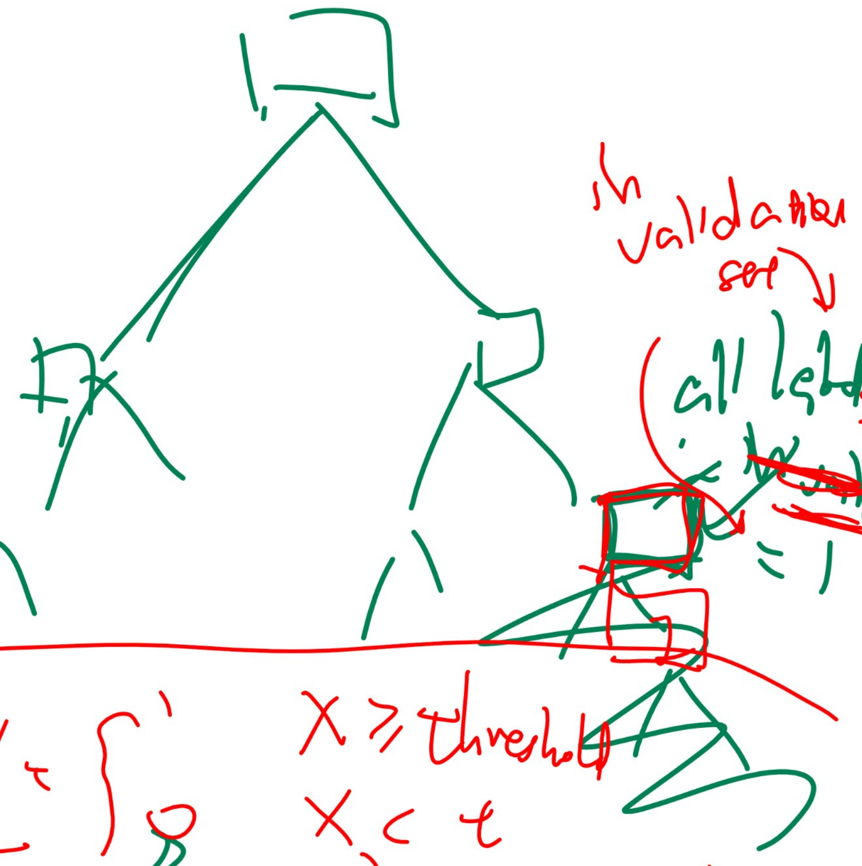
regularization.

tree

train

train

validation



$X \geq$ threshold

$X < t$

binary split

if continuous feature

$$H(Y|X) \rightarrow \sum_{i=1}^{10} p_{x_i} H(Y|X=x_i)$$

Bagging

~~Discussion~~

LTU
Logistic ⇒ NN

SVM ⟹ kernel

DTree ⇒ Random Forest

- When training the decision trees on the smaller training sets, only a ran~~dom subset of the features~~ are used. The decision trees are created without pruning.
- This algorithm is called random forest.

Bootstrap
Aggregating

K Nearest Neighbor

Description

- Given a new instance, find the K instances in the training set that are the closest.
- Predict the label of the new instance by the majority of the labels of the K instances.

Distance Function

Definition

- Many distance functions can be used in place of the Euclidean distance.

p-norm

$$\rho\left(x, x'\right) = \left\|x - x'\right\|_2 = \sqrt{\sum_{j=1}^{m} \left(x_j - x'_j\right)^2}$$

L1 m

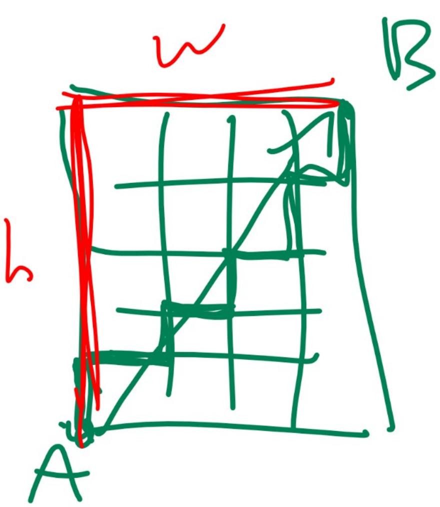
- An example is Manhattan distance.

KNN

Cityblock

$$\rho\left(x, x'\right) = \sum_{j=1}^{m} \left|x_j - x'_j\right|$$

Taxicab



1 Nearest Neighbor

Quiz

test item

- Spring 2018 Midterm Q7
- Find the 1 Nearest Neighbor label for $\begin{bmatrix}3\\6\end{bmatrix}$ using Manhattan distance.

x_1	1	1	3	5	2
x_2	1	7	3	4	5
y	0	1	1	0	0

~~6~~ 3 3 4 2

7

- A: 0
- B: 1

3 Nearest Neighbor

Quiz

Q2

K = 3

~~2NN~~

K

- Find the $\boxed{\text{3 Nearest Neighbor}}$ label for $\boxed{\begin{bmatrix} 3 \\ 6 \end{bmatrix}}$ using Manhattan distance.

x_1	1	1	3	5	2
x_2	1	7	3	4	5
y	0	1	1	0	0

7 3 3 4 2 ← distances

2, 3 1

$\frac{1}{3}$ $\frac{1}{3}$ $\frac{1}{2}$

distance

vote

weighted KNN

- A: 0
- B: 1

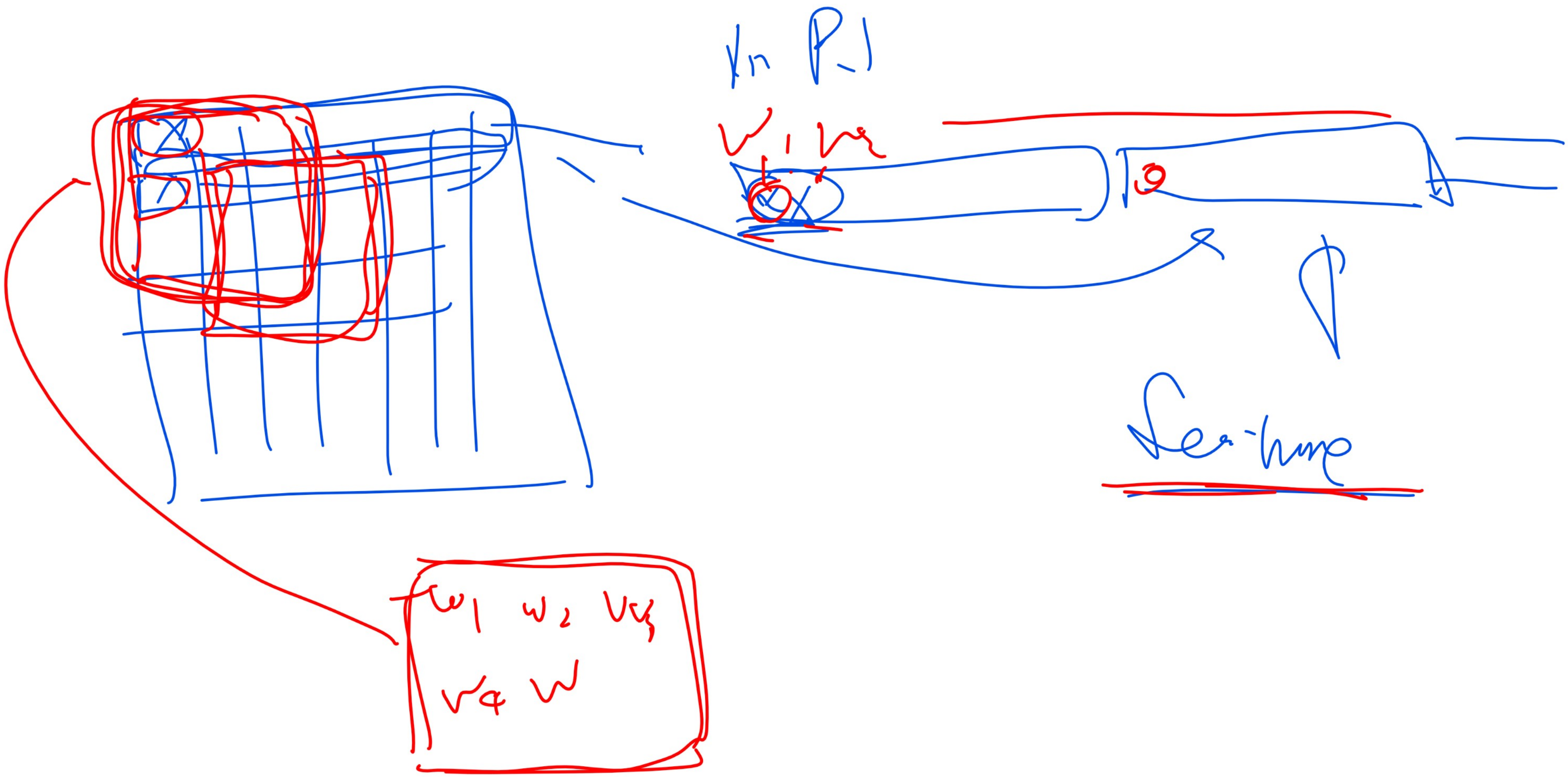
$$d\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}3\\6\end{pmatrix}\right) = |1-3| + |1-6| = 2+5 = 7$$

Computer Vision Examples

Motivation

Image Features Diagram

Motivation



One Dimensional Convolution

Definition

- The convolution of a vector $x = (x_1, x_2, ..., x_m)$ with a filter $w = (w_{-k}, w_{-k+1}, ... w_{k-1}, w_k)$ is:

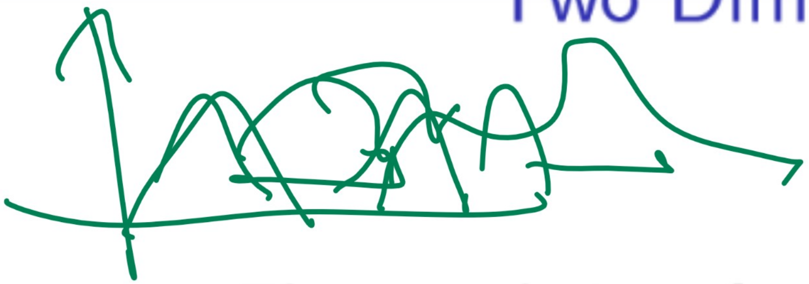
$$a = (a_1, a_2, ..., a_m) = x * w$$

$$a_j = \sum_{t=-k}^{k} w_t x_{j-t}, j = 1, 2, ..., m$$

$\approx w^T x$

- w is also called a kernel (different from the kernel for SVMs).
- The elements that do not exist are assumed to be 0.

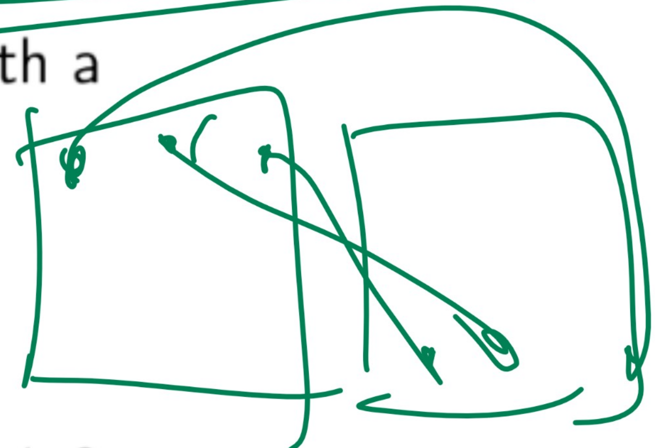
Two Dimensional Convolution



$\int f(t)\, g(s-t)\, dt$

Definition

- The convolution of an $m \times m$ matrix X with a $(2k+1) \times (2k+1)$ filter W is:



$$A = X * W$$

$$A_{j,j'} = \boxed{\sum_{s=-k}^{k} \sum_{t=-k}^{k} W_{s,t} X_{j-s,j'-t}}, j, j' = 1, 2, ..., m$$

- The matrix W is indexed by (s, t) for $s = -k, -k+1, ..., k-1, k$ and $t = -k, -k+1, ..., k-1, k$.
- The elements that do not exist are assumed to be 0.

Convolution Diagram and Demo

Definition

Image Gradient

Definition

- The gradient of an image is defined as the change in pixel intensity due to the change in the location of the pixel.

$$\frac{\partial I(s,t)}{\partial s} \approx \boxed{\frac{I\left(s+\frac{\varepsilon}{2},t\right)-I\left(s-\frac{\varepsilon}{2},t\right)}{\varepsilon}}, \varepsilon = 1$$

$$\frac{\partial I(s,t)}{\partial t} \approx \frac{I\left(s,t+\frac{\varepsilon}{2}\right)-I\left(s,t-\frac{\varepsilon}{2}\right)}{\varepsilon}, \varepsilon = 1$$

Image Derivative Filters

Definition

- The gradient can be computed using convolution with the following filters.

$$w_x = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}, w_y = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Sobel Filter

Definition

- The Sobel filters also are used to approximate the gradient of an image.

$$W_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, W_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

Gradient of Images

Definition

- The gradient of an image I is $(\nabla_x I, \nabla_y I)$.

$$\nabla_x I = W_x * I, \nabla_y I = W_y * I$$

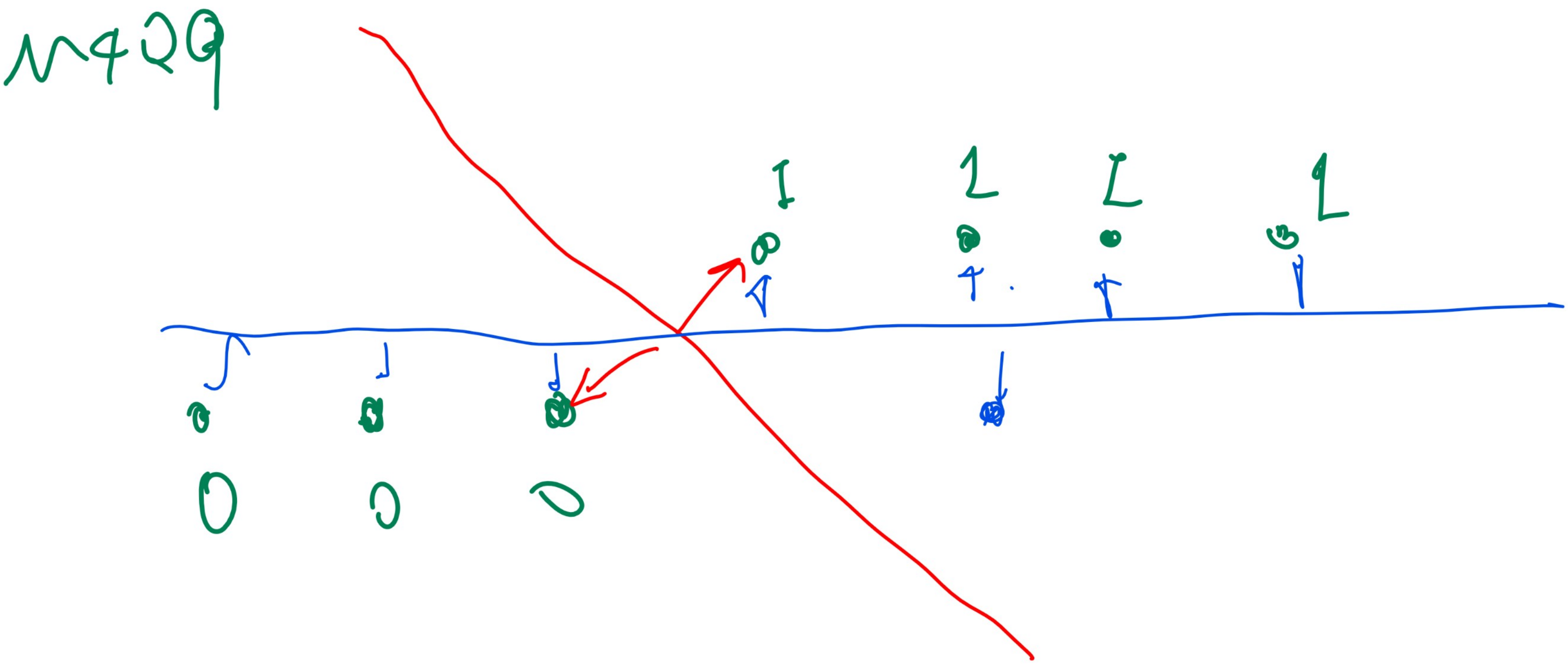
- The gradient magnitude is G and gradient direction Θ are the following.

$$G = \sqrt{\nabla_x^2 + \nabla_y^2}$$

$$\Theta = \arctan\left(\frac{\nabla_y}{\nabla_x}\right)$$

Gradient of Images Demo

Definition



Convolution Example

Quiz

- 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

$0 \cdot 0 + 0 \cdot 0 + -1(1) + 0 \cdot 1$

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

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

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

image $= \begin{bmatrix}0_0 & 0_1 & 0_0\\1_0 & 1_0 & 1_0\\1_0 & 1_{-1} & 1_0\end{bmatrix}$

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

Gradient Example

Quiz

Convolution Example 1

Quiz

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

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

$\times$ sobel

- 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 2

Quiz

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

- 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 3

Quiz

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

- 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

What is the gradient magnitude for the center cell?

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

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

Convolution Example 5

Quiz

What is the gradient direction bin for the center cell?

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

- A: - π, B: - $\frac{\pi}{2}$, C: 0, D: $\frac{\pi}{2}$, E: π

SIFT

Discussion

- Scale Invariant Feature Transform (SIFT) features are features that are invariant to changes in the location, scale, orientation, and lighting of the pixels.

Histogram Binning Diagram

Discussion

HOG

Discussion

- Histogram of Oriented Gradients features is similar to SIFT but does not use dominant orientations.

Matching vs Classification Diagram

Discussion