Deep Learning for Computer Vision

Lecture 1 - Overview

Dr. José Ramón Iglesias
DSP-ASIC BUILDER GROUP
Director Semillero TRIAC
Ingenieria Electronica
Universidad Popular del Cesar

Today’s agenda

- A brief history of computer vision
- In overview

Today’s agenda

- A brief history of computer vision
- In overview

In overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

Deep Learning Basics

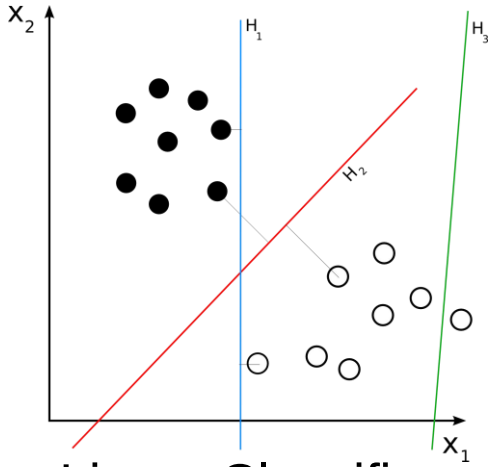
- Image Classification: A core task in Computer Vision

→ cat

Deep Learning Basics

- Image Classification: A core task in Computer Vision

cat

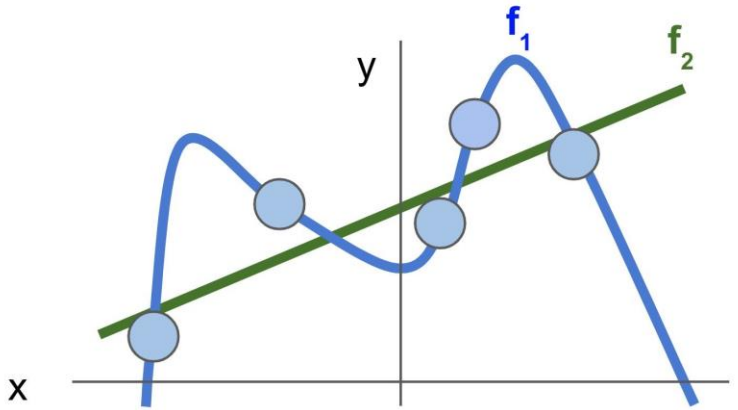


Linear Classifier

Deep Learning Basics

- Image Classification: A core task in Computer Vision

cat

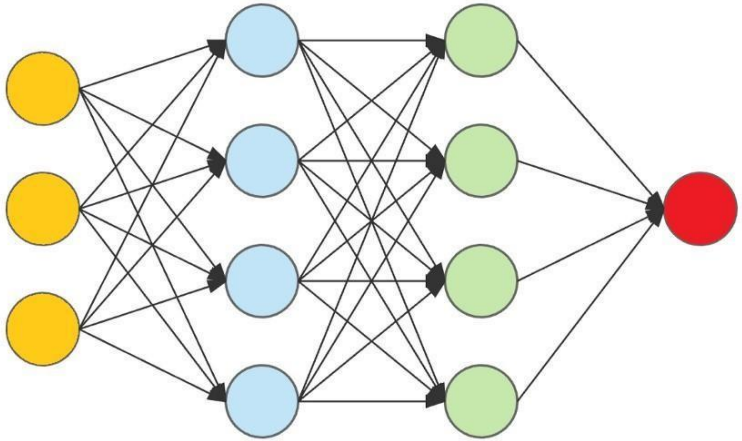


Regularization & Optimization

Deep Learning Basics

- Image Classification: A core task in Computer Vision

cat



Neural Networks

CS231n overview

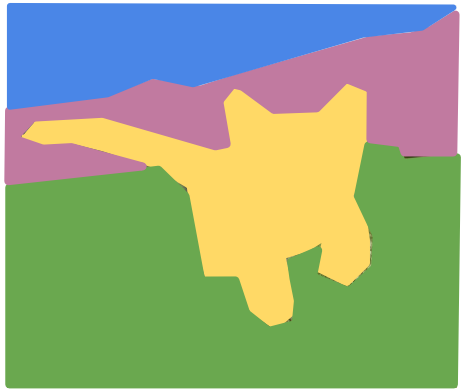
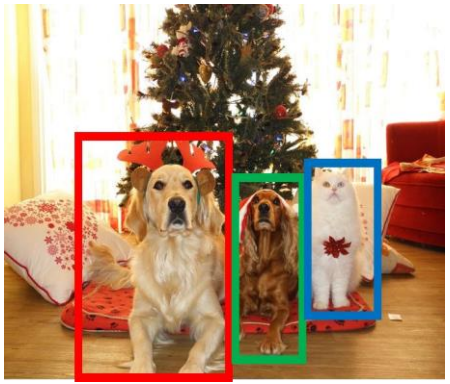
- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

In overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

Perceiving and Understanding the Visual World

Tasks Beyond Image Classification

Classification	Semantic Segmentation	Object Detection	Instance Segmentation
			
CAT	**GRASS**, **CAT**, **TREE**, **SKY**	**DOG**, **DOG**, **CAT**	**DOG**, **DOG**, **CAT**
No spatial extent	No objects, just pixels	Multiple Object	

Tasks Beyond Image Classification

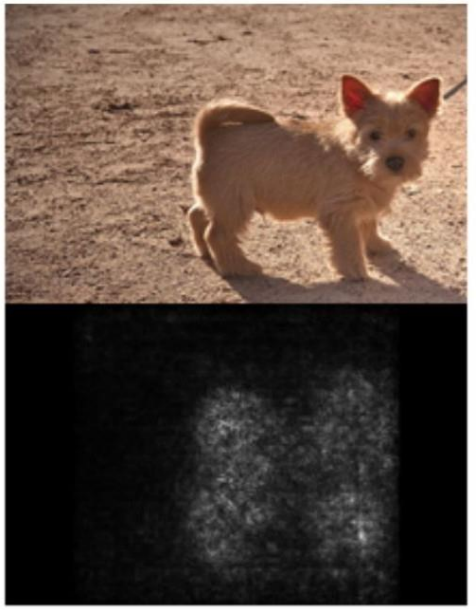
Video Classification

Running?
Jumping?

Multimodal Video Understanding

Introducción

Visualization & Understanding



Models Beyond Multi-Layer Perceptron

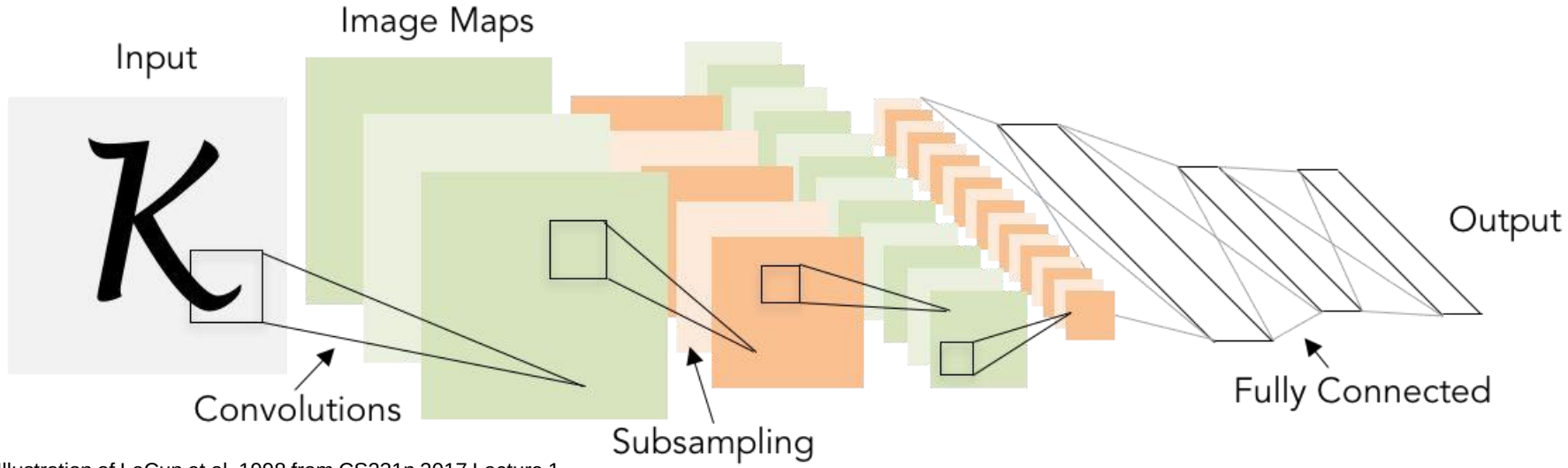
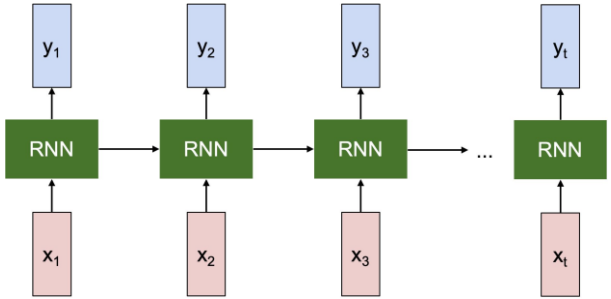
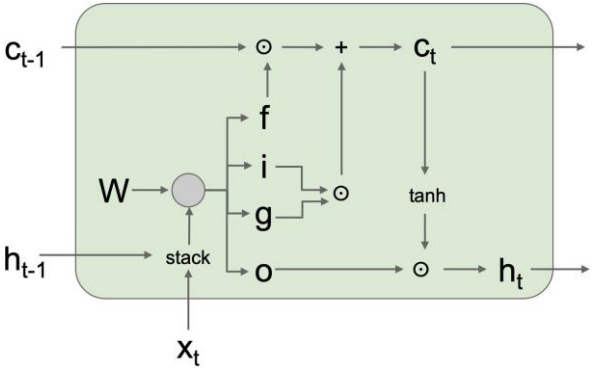


Illustration of LeCun et al. 1998 from CS231n 2017 Lecture 1

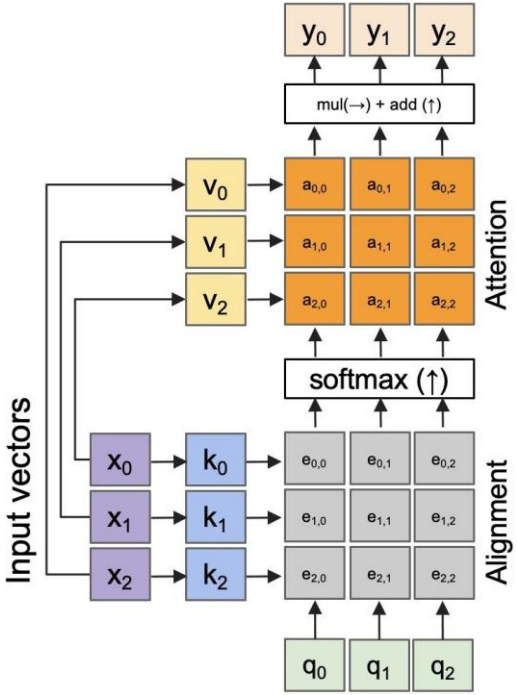
Convolutional neural network

Models Beyond Multi-Layer Perceptron





Recurrent neural network



Attention mechanism / Transformers

In overview

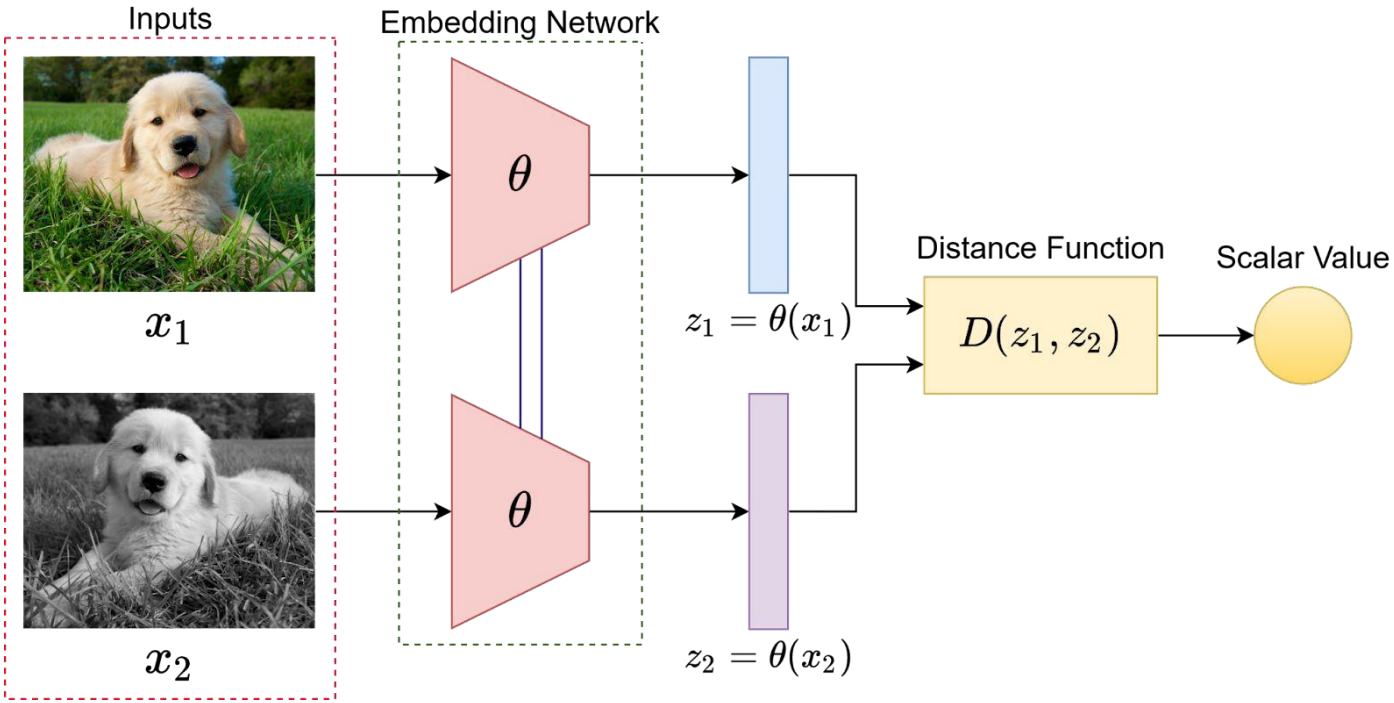
- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

In overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

Beyond 2D Recognition

Beyond 2D Recognition: Self-supervised Learning

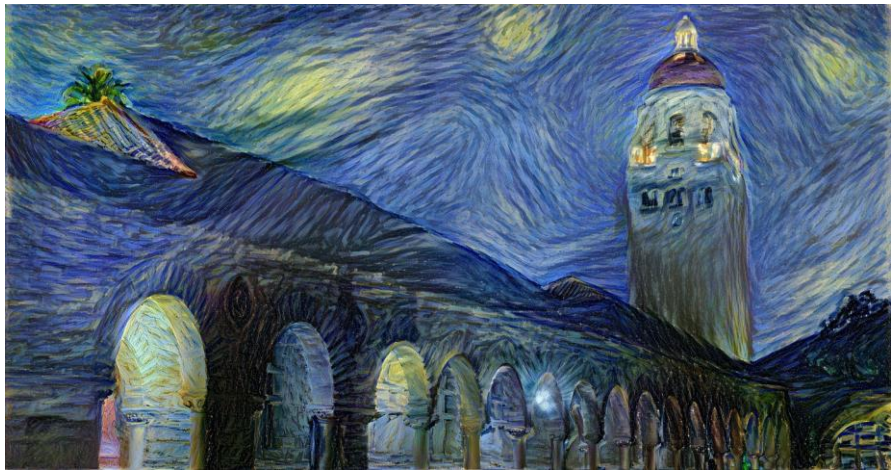


Beyond 2D Recognition: Generative Modeling

“Teddy bears working on new AI research underwater with 1990s technology”

DALL-E 2

Beyond 2D Recognition: Generative Modeling

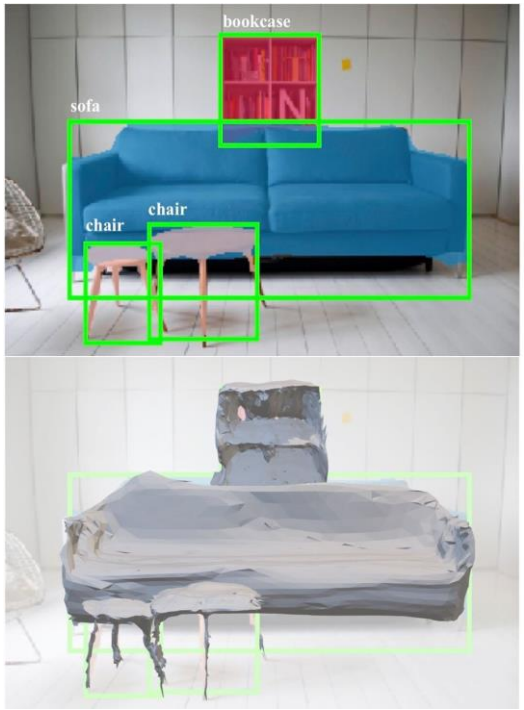


Style Transfer

Beyond 2D Recognition: 3D Vision

Choy et al., 3D-R2N2: Recurrent Reconstruction Neural Network (2016)

Zhou et al., 3D Shape Generation and Completion through Point-Voxel Diffusion (2021)



Gkioxari et al., "Mesh R-CNN", ICCV 2019

Beyond 2D Recognition: Embodied Intelligence

Li et al., BEHAVIOR-1K: A Benchmark for Embodied AI with 1,000 Everyday Activities and Realistic Simulation (2022)

Mandlekar and Xu et al., Learning to Generalize Across Long-Horizon Tasks from Human Demonstrations (2020)

In overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

In overview

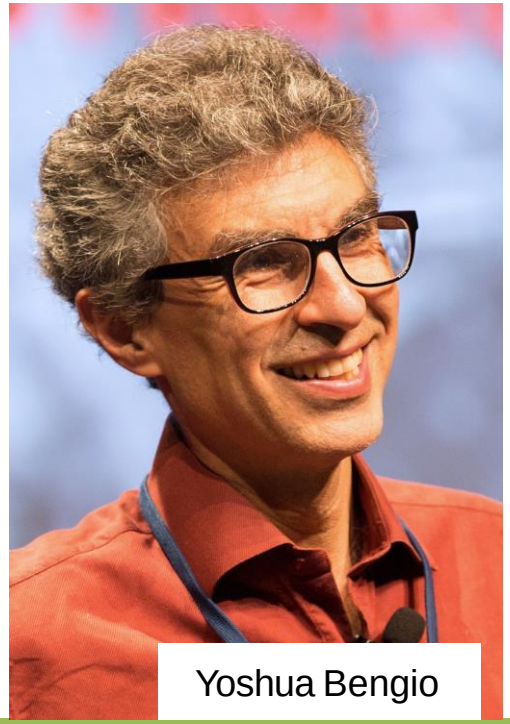
- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

2018 Turing Award for deep learning

most prestigious technical award, is given for major contributions of lasting importance to computing.

Jeffrey Hinton

This image. CC0 public domain



Yoshua Bengio

This image. CC0 public domain

Yann LeCun

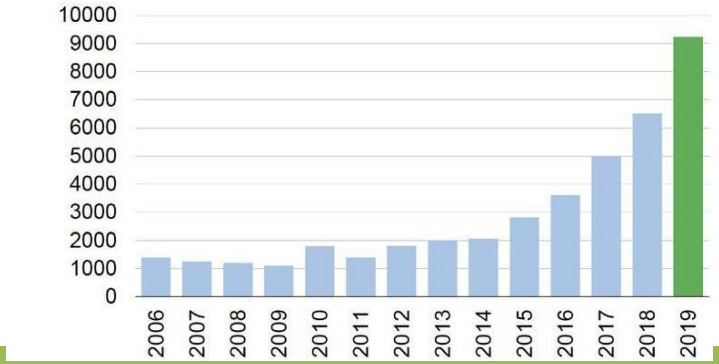
This image. CC0 public domain

IEEE PAMI Longuet-Higgins Prize

Award recognizes ONE Computer Vision paper from **ten years ago** with **significant impact on computer vision** research.

At CVPR 2019, it was awarded to the 2009 original ImageNet paper

CVPR Attendance Trend



JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

>9k submissions, 2,360 accepted papers

Optional textbook resources

- *Deep Learning*
  - by Goodfellow, Bengio, and Courville
  - Here is a free version
- Mathematics of deep learning
  - Chapters 5, 6 7 are useful to understand vector calculus and continuous optimization
  - Free online version
- Dive into deep learning
  - An interactive deep learning book with code, math, and discussions, based on the NumPy interface.
  - Free online version

Learning objectives

Formalize computer vision applications into tasks

- Formalize inputs and outputs for vision-related problems
- Understand what data and computational requirements you need to train a model

Develop and train vision models

- Learn to code, debug, and train convolutional neural networks.
- Learn how to use software frameworks like PyTorch and TensorFlow

Gain an understanding of where the field is and where it is headed

- What new research has come out in the last 0-5 years?
- What are open research challenges?
- What ethical and societal considerations should we consider before deployment?

Why should you take this class?

Become a vision researcher (an incomplete list of conferences)

- Get involved with vision research at Stanford: apply using this form.
- CVPR 2022 conference
- ICCV 2021 conference

Become a vision engineer in industry (an incomplete list of industry teams)

- Perception team at Google AI, Vision at Google Cloud
- Vision at Meta AI
- Vision at Amazon AWS
- Nvidia, Tesla, Apple, Salesforce, ……

General interest

Deep Learning for Computer Vision

- Deep Learning Basics (Lecture 2 – 4)
- Perceiving and Understanding the Visual World (Lecture 5 – 12)
- Reconstructing and Interacting with the Visual World (Lecture 13 – 16)
- Human-Centered Artificial Intelligence (Lecture 17 – 18)

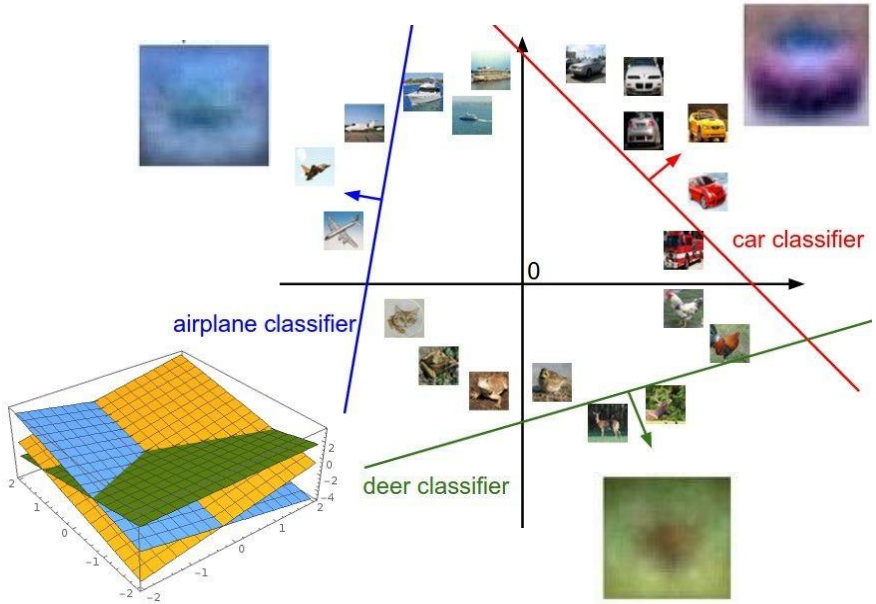
Syllabus

Deep Learning Basics	Convolutional Neural Networks	Computer Vision Applications
Data-driven learning	Convolutions	RNNs / Attention / Transformers
Linear classification & kNN	PyTorch / TensorFlow	Image captioning
Loss functions	Activation functions	Object detection and segmentation
Optimization	Batch normalization	Style transfer
Backpropagation	Transfer learning	Video understanding
Multi-layer perceptrons	Data augmentation	Generative models
Neural Networks	Momentum / RMSProp / Adam	Self-supervised learning
	Architecture design	3D vision
		Robot learning
		Human-centered AI
		Fairness & ethics

Next time: Image classification with Linear Classifiers

k- nearest neighbor

Linear classification



Introducción

Universidad Popular del Cesar