Introduction to Machine Learning and Deep Learning

Овчинников Илья

[DASHA | AI]

Летняя международная XXXI молодежная Школа-конференция по параллельному программированию

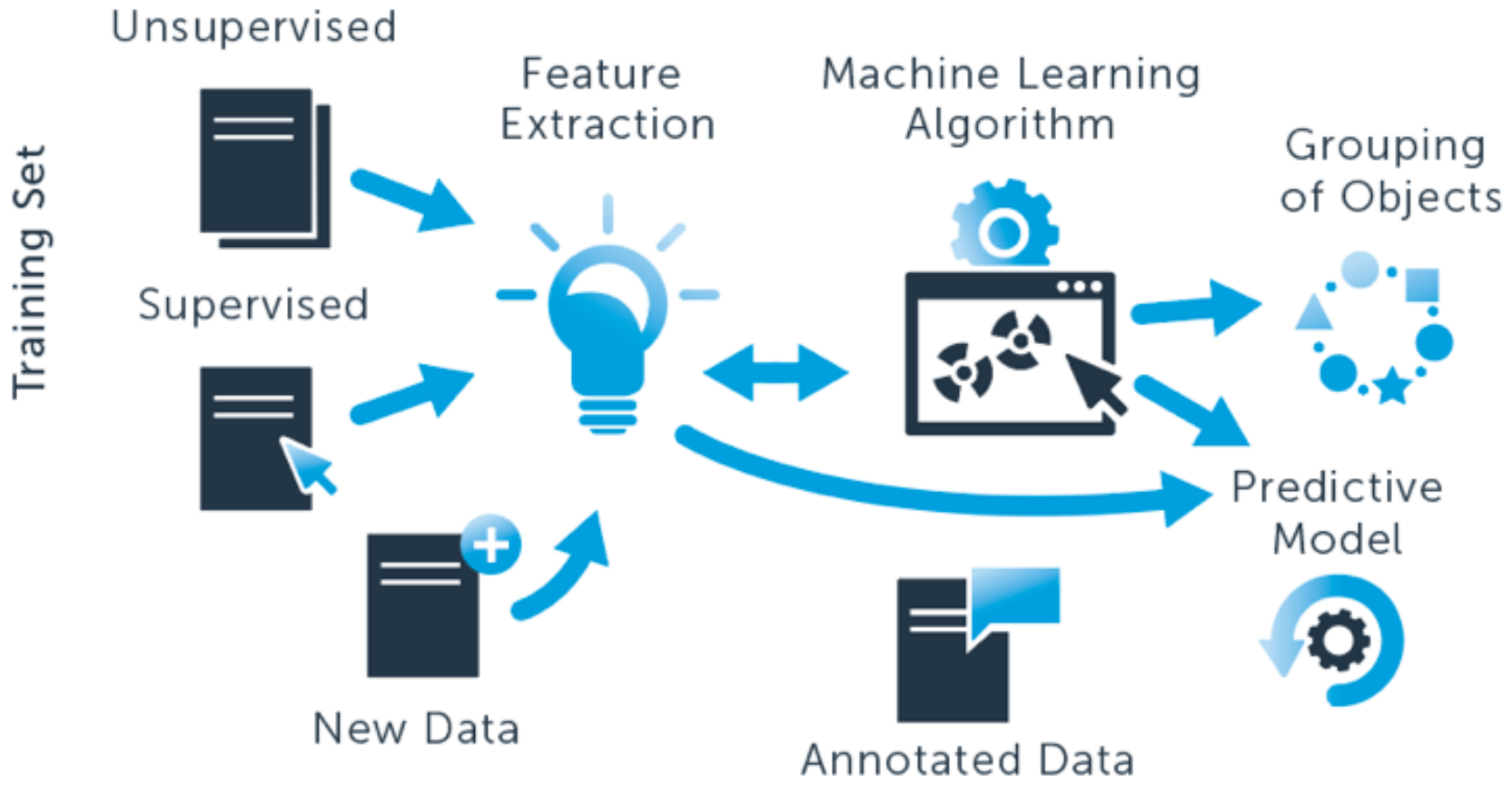
Новосибирск

10 июля, 2018

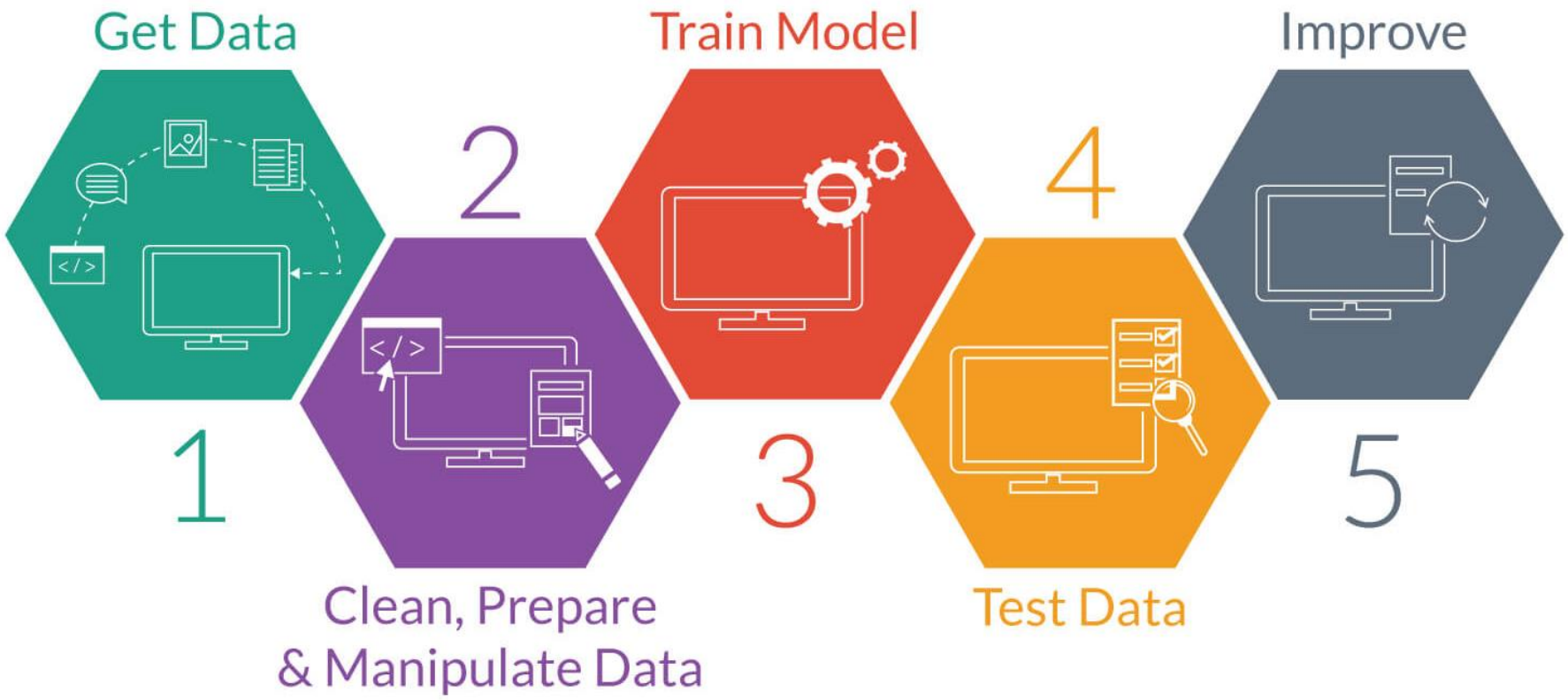
Почему Machine Learning?

- Обработка больших данных
- Вычислительные мощности
- Выделение скрытых зависимостей
- Сложные математические модели
- Практическое применение

Machine Learning



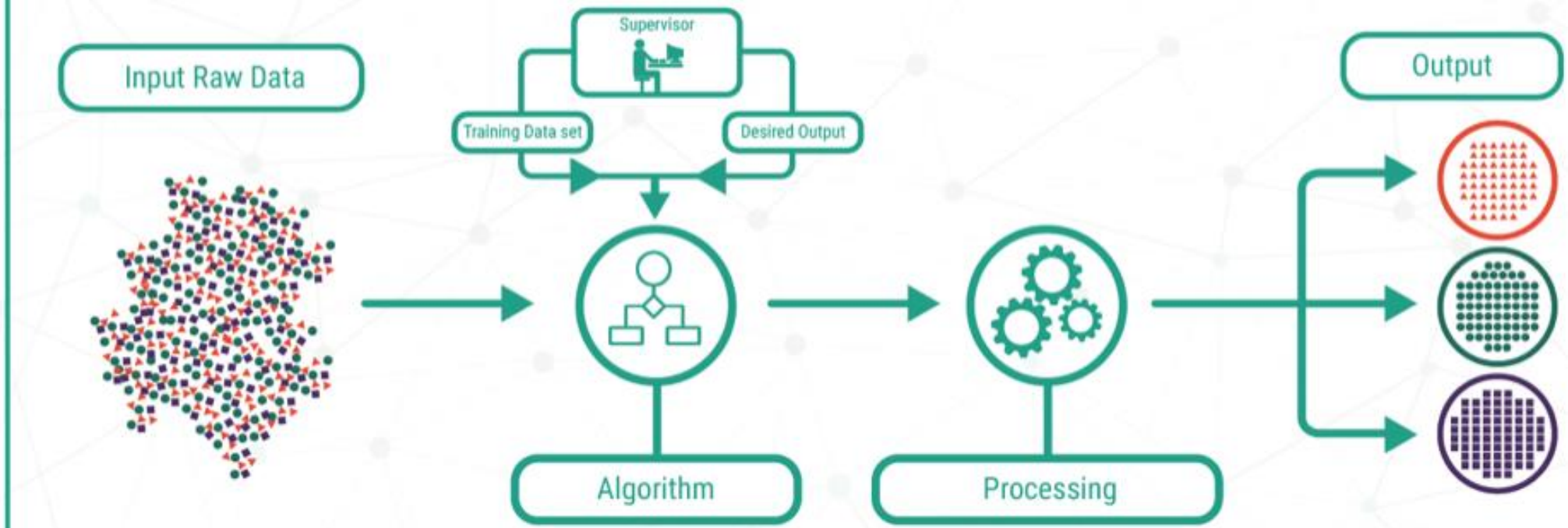
Общий алгоритм машинного обучения

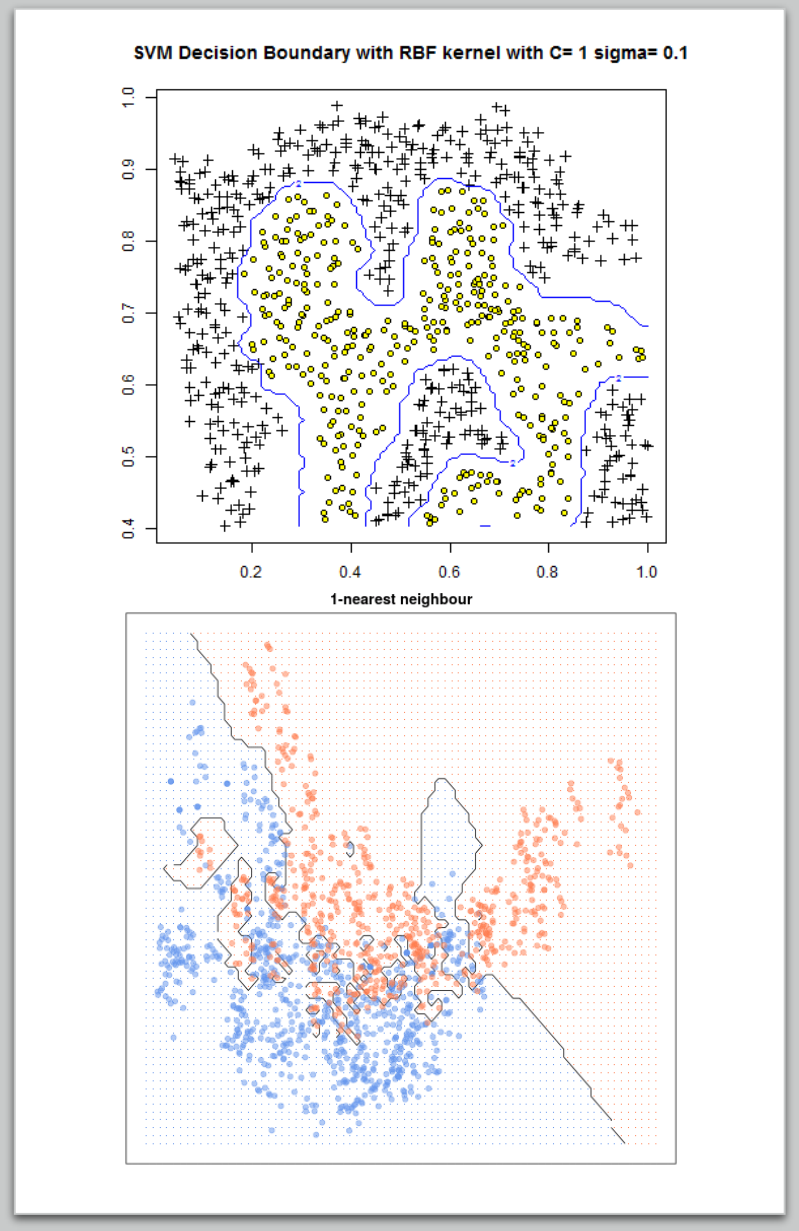


Machine Learning

- Supervised
  - Размеченные данные $\mathrm{D} = \{(X, y)\}$
  - Построить функцию предсказания $f(X) = y$
  - Классификация: y - дискретна
  - Регрессия: y - непрерывна
- Unsupervised
  - Только неразмеченные данные $\mathrm{D} = \{(X)\}$
  - Построить восстанавливающую функцию $f(X) = y$
  - Кластеризация
- Semi-supervised
  - Размеченные + не размеченные данные $\mathrm{D} = \{(X, y), (X)\} \rightarrow f(X) = y$

SUPERVISED LEARNING





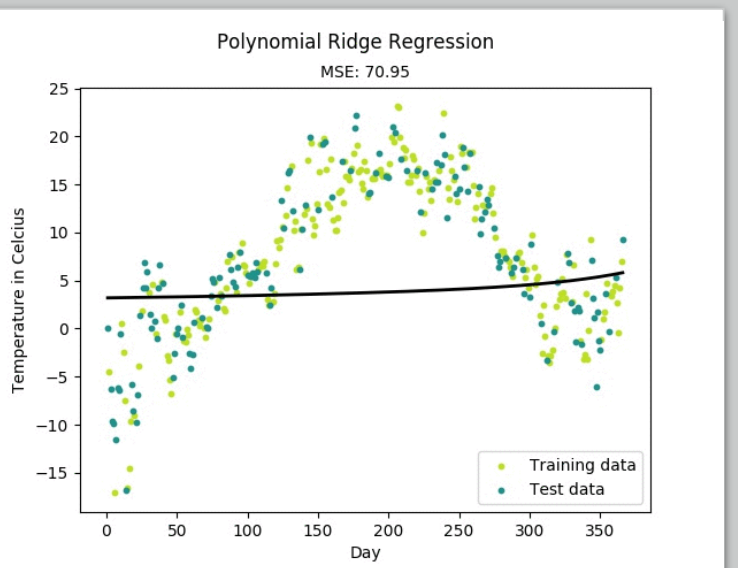
Классификация

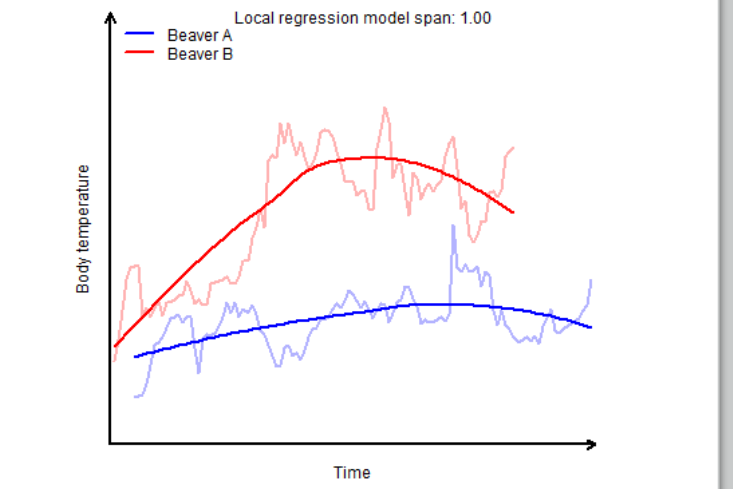
Примеры:

- Распознавание объектов на изображениях
- Credit risk management
- Выявление заболеваний у пациентов
- Определение смысла текста
- Выявление мошенников
- Эмоциональный окрас текста
- Рекомендательные системы

Алгоритмы:

- Nearest Neighbors
- Naive Bayes
- Decision Trees
- Support Vector Machines
- Логистическая регрессия
- XGBoost
- Neural Networks





Регрессия

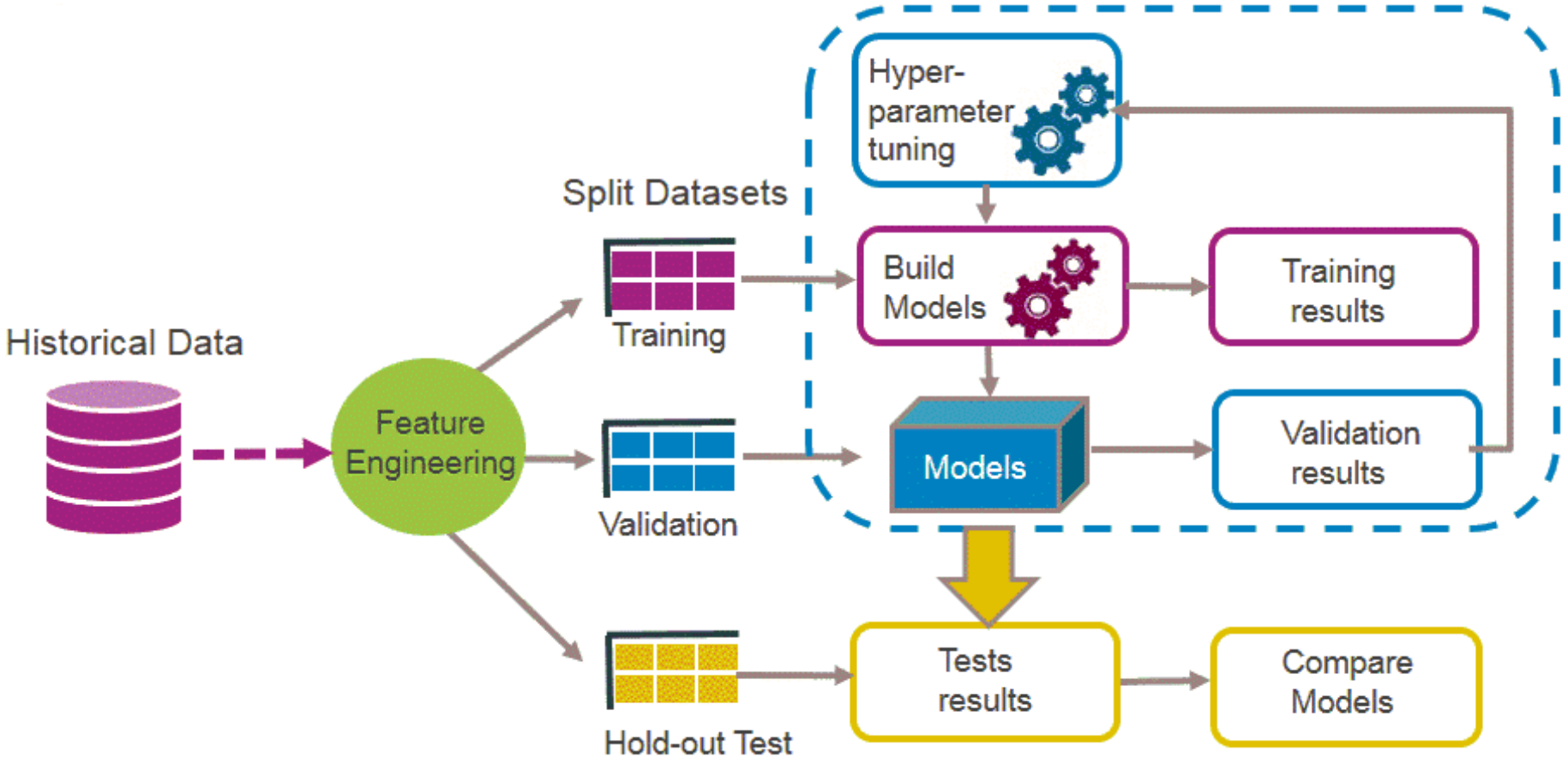
Примеры:

- Предсказание рыночной стоимости квартир
- Определения качества продуктов
- Оценка успеваемости студентов
- Предсказание цен на биржевые акции
- Оценка популярности постов в социальных сетях

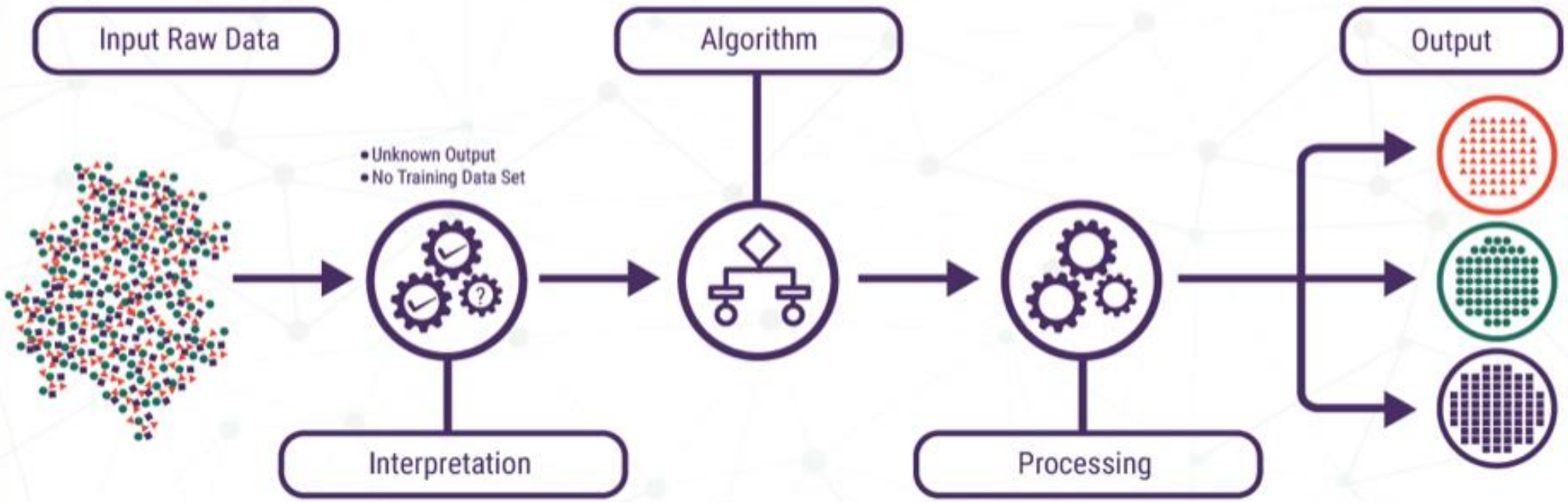
Алгоритмы:

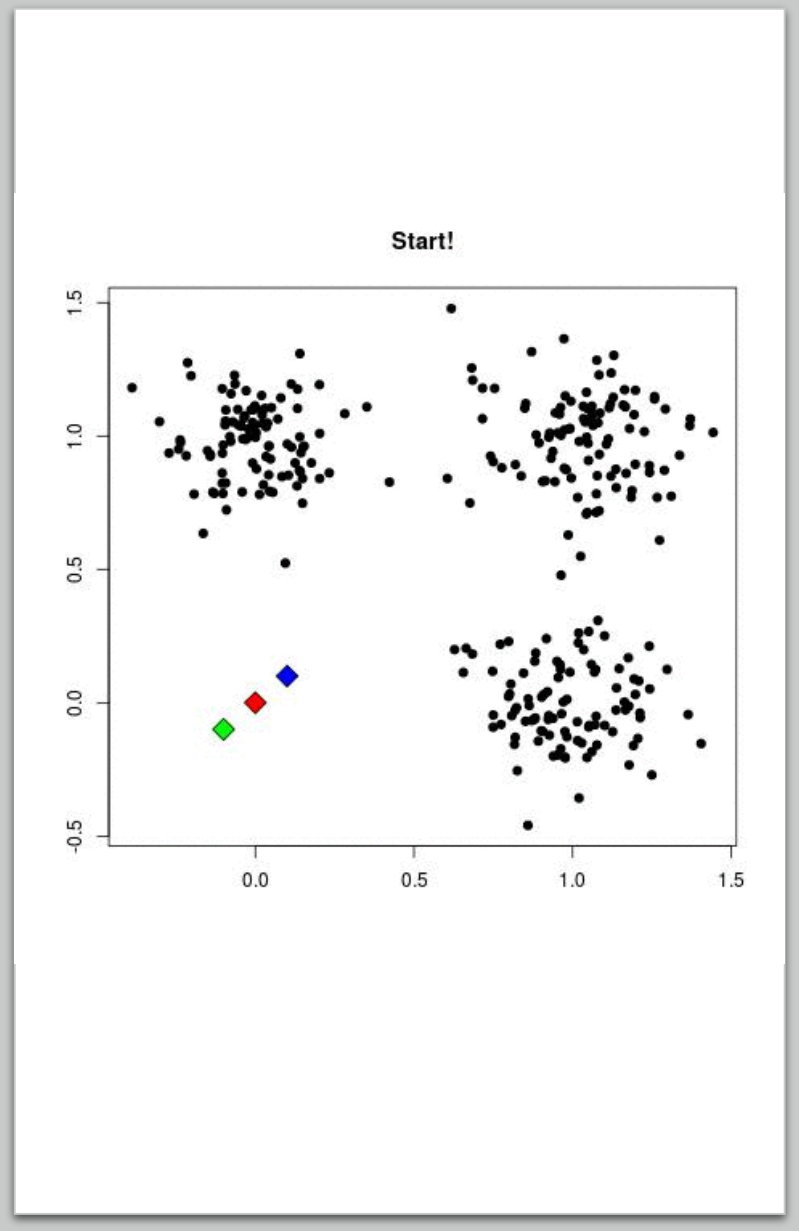
- Метод наименьших квадратов
- Ridge и Lasso регрессия
- Логистическая регрессия
- Support Vector Machines
- Decision Trees
- Neural networks
- Ensemble methods
- XGBoost

Построение предиктивной модели



UNSUPERVISED LEARNING





Кластеризация

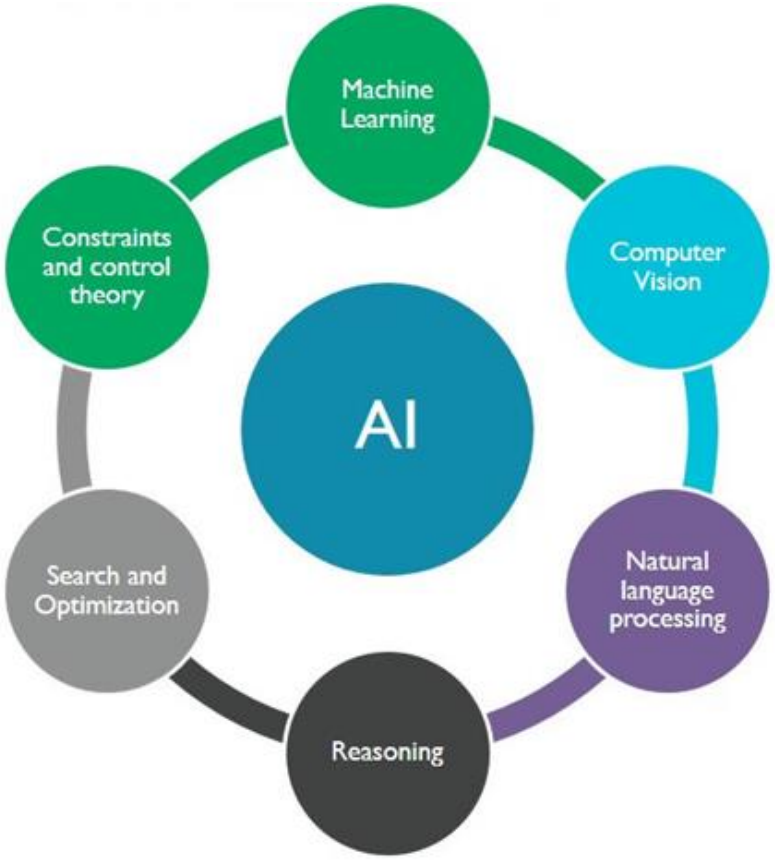
Примеры:

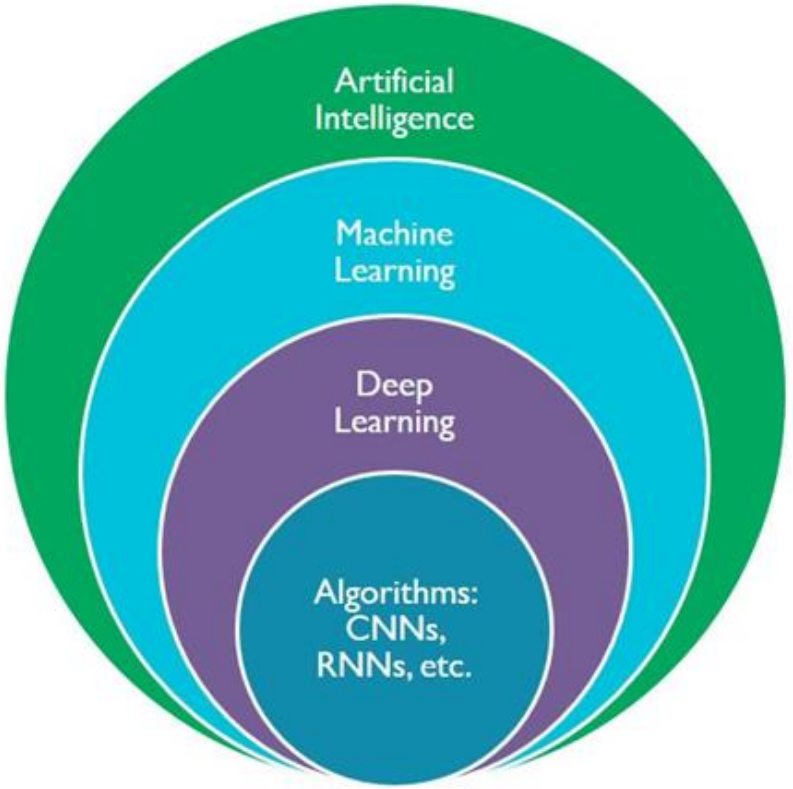
- Выделение групп покупателей
- Рекомендательные системы
- Группировка поездок такси
- Выделение активности пользователей

Алгоритмы:

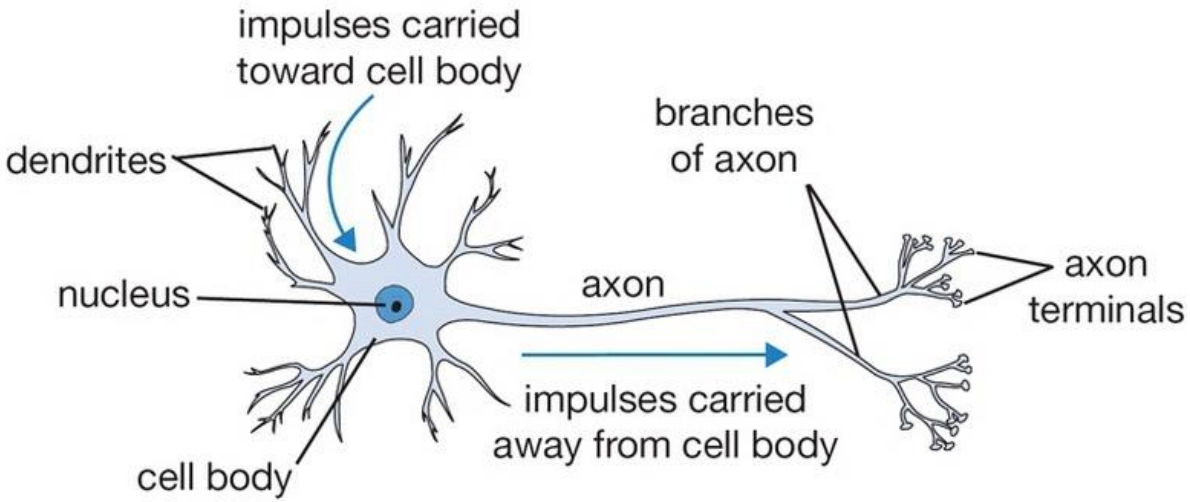
- K-means
- Affinity Propagation
- Hierarchical clustering
- DBSCAN

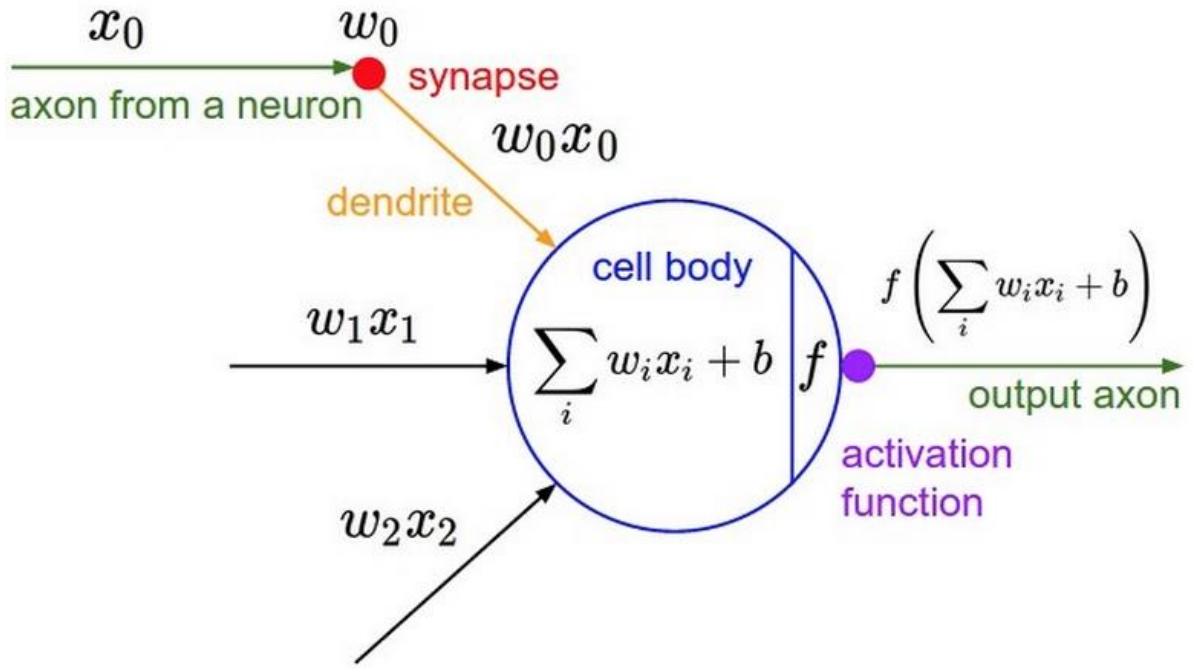
Artificial Intelligence

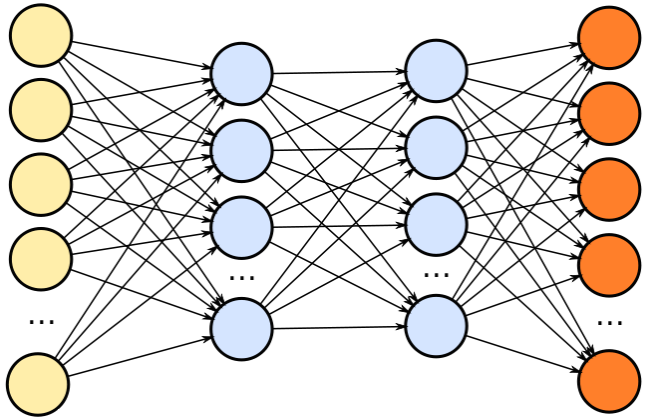


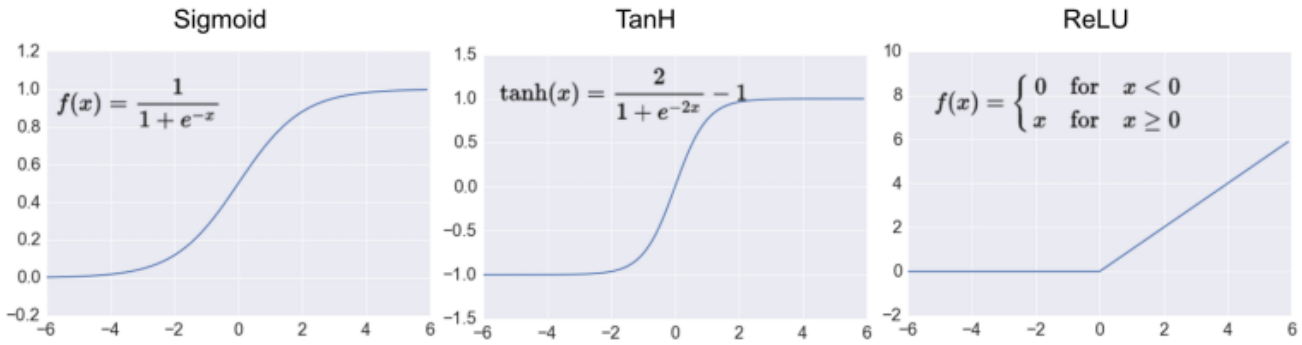


Artificial Neural Networks

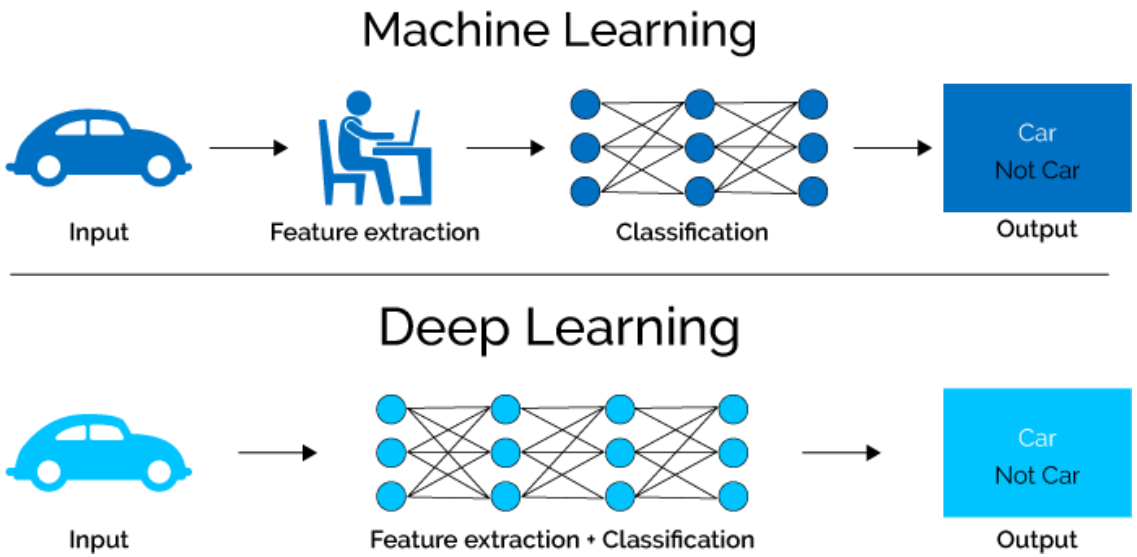


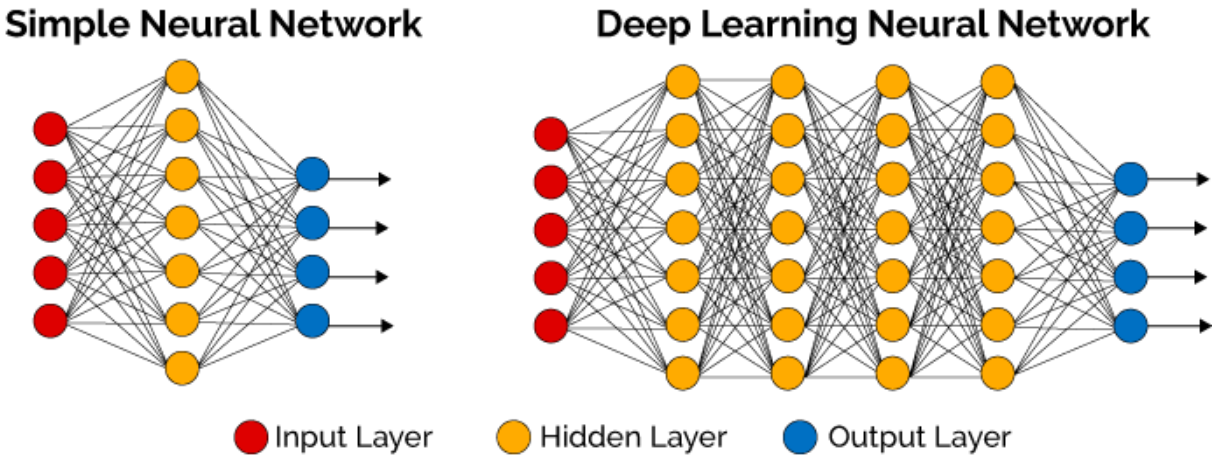




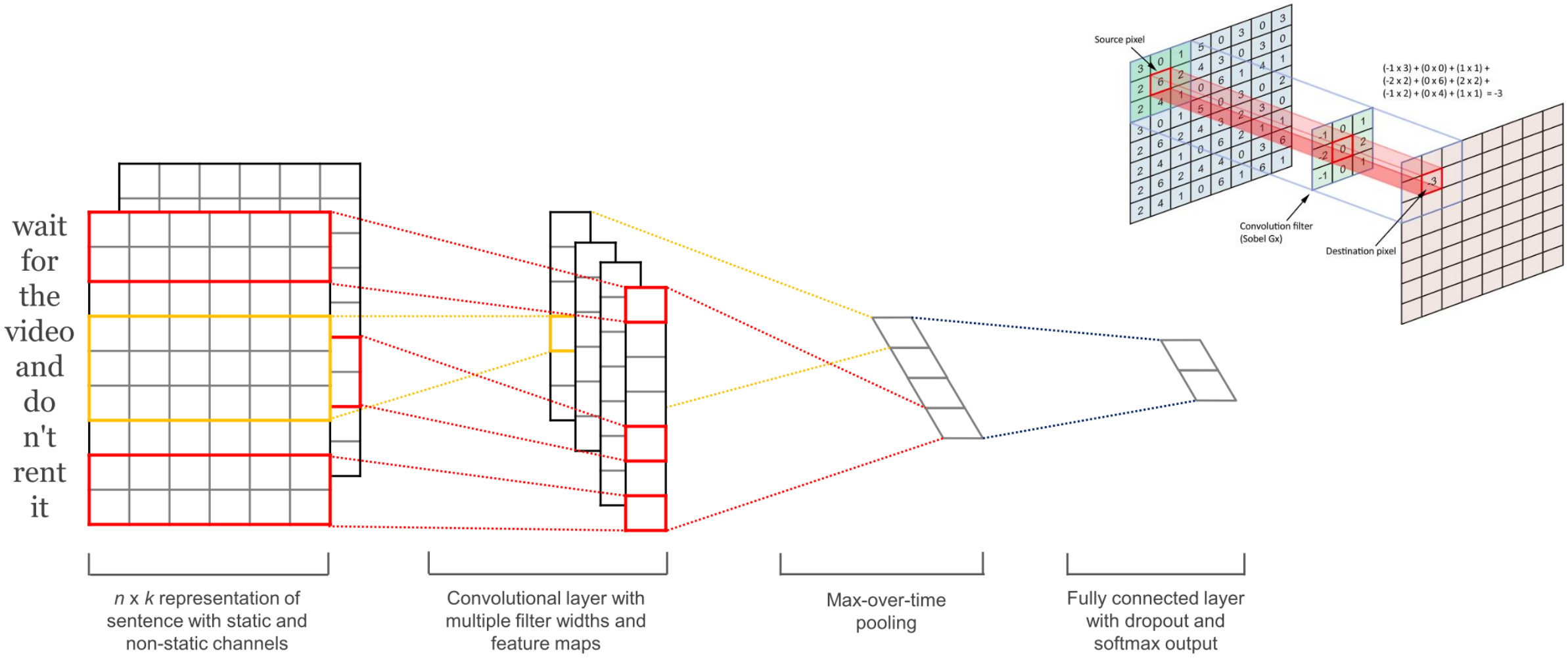


Deep Learning

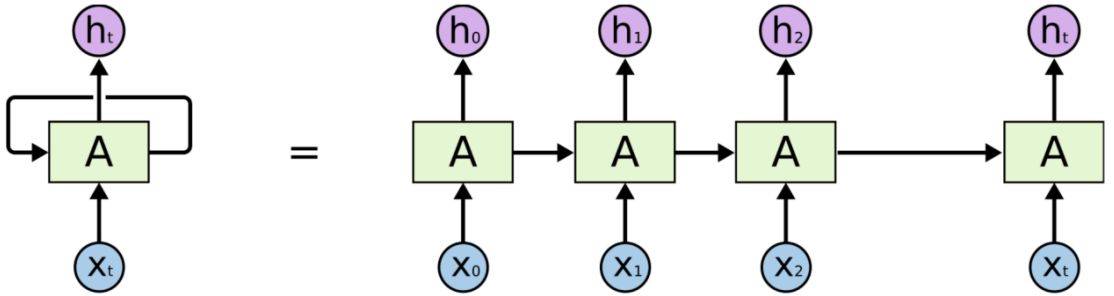


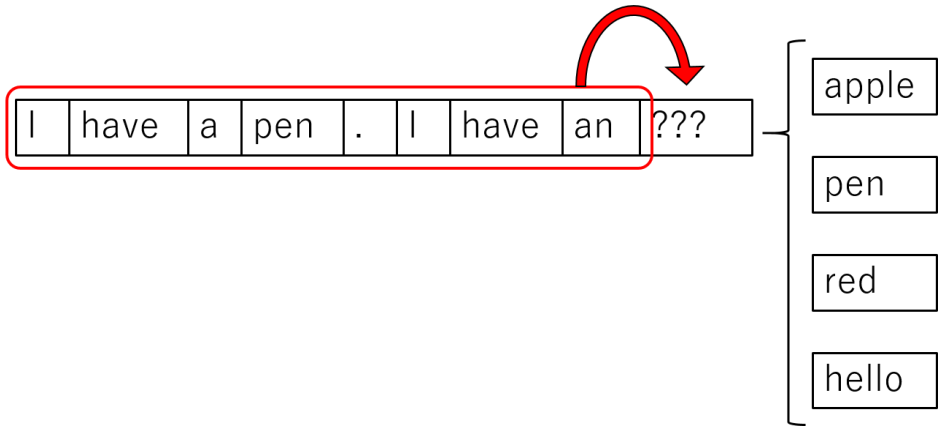


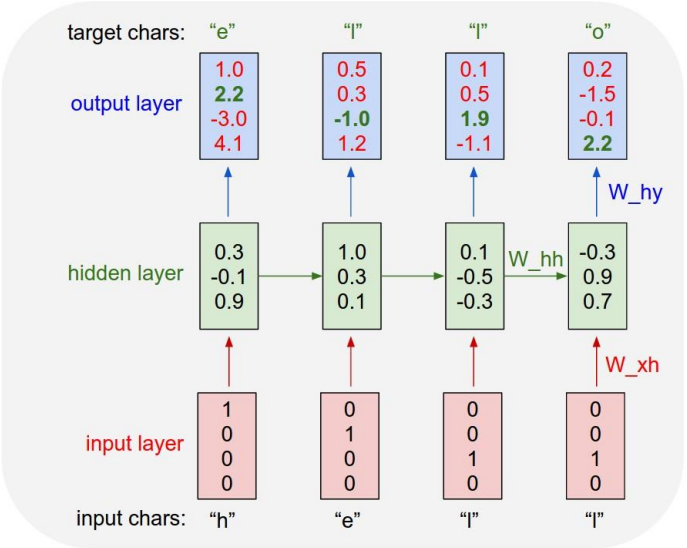
Convolutional Neural Networks



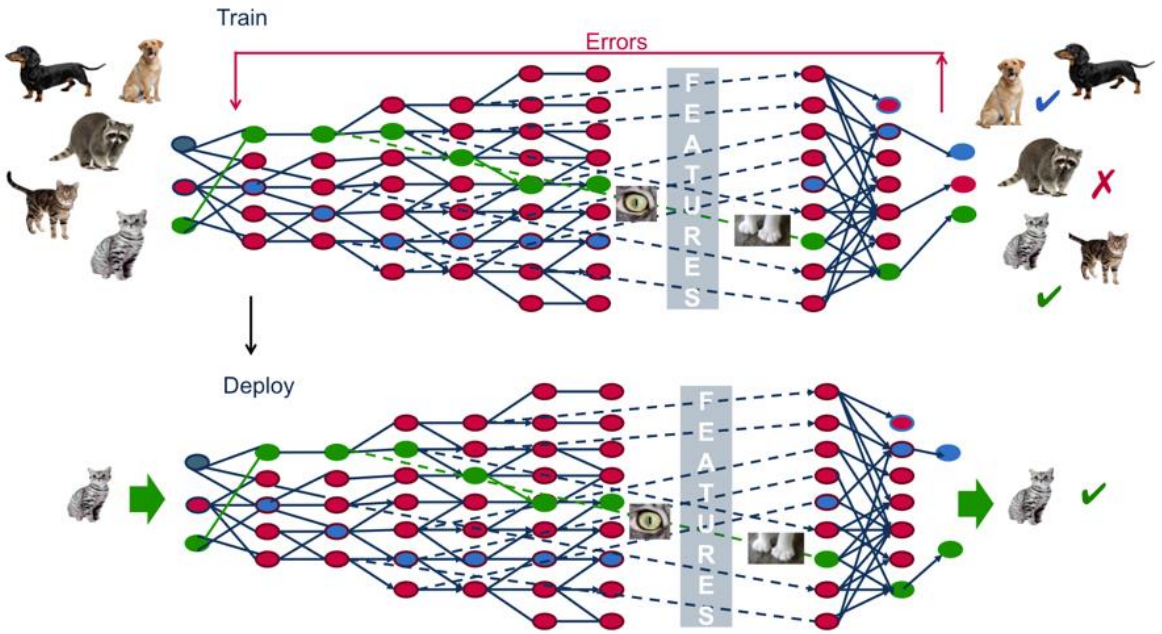
Recurrent Neural Networks

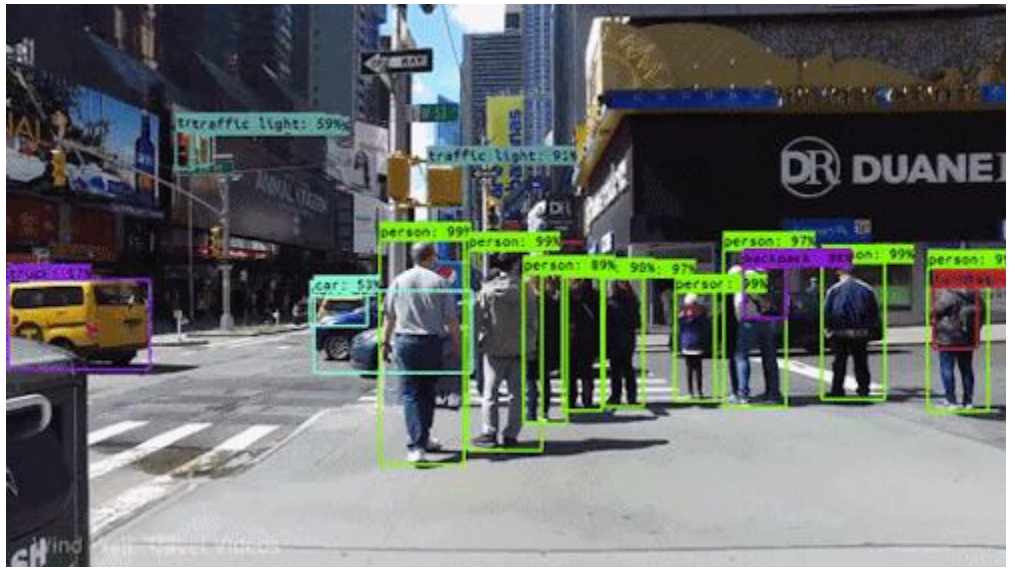


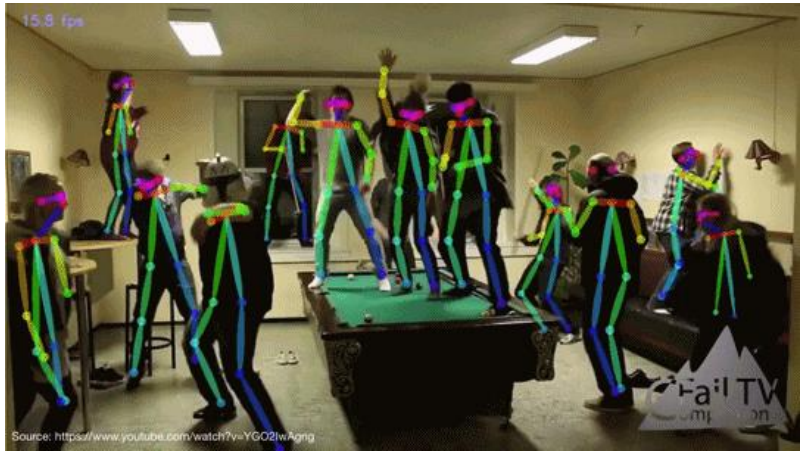




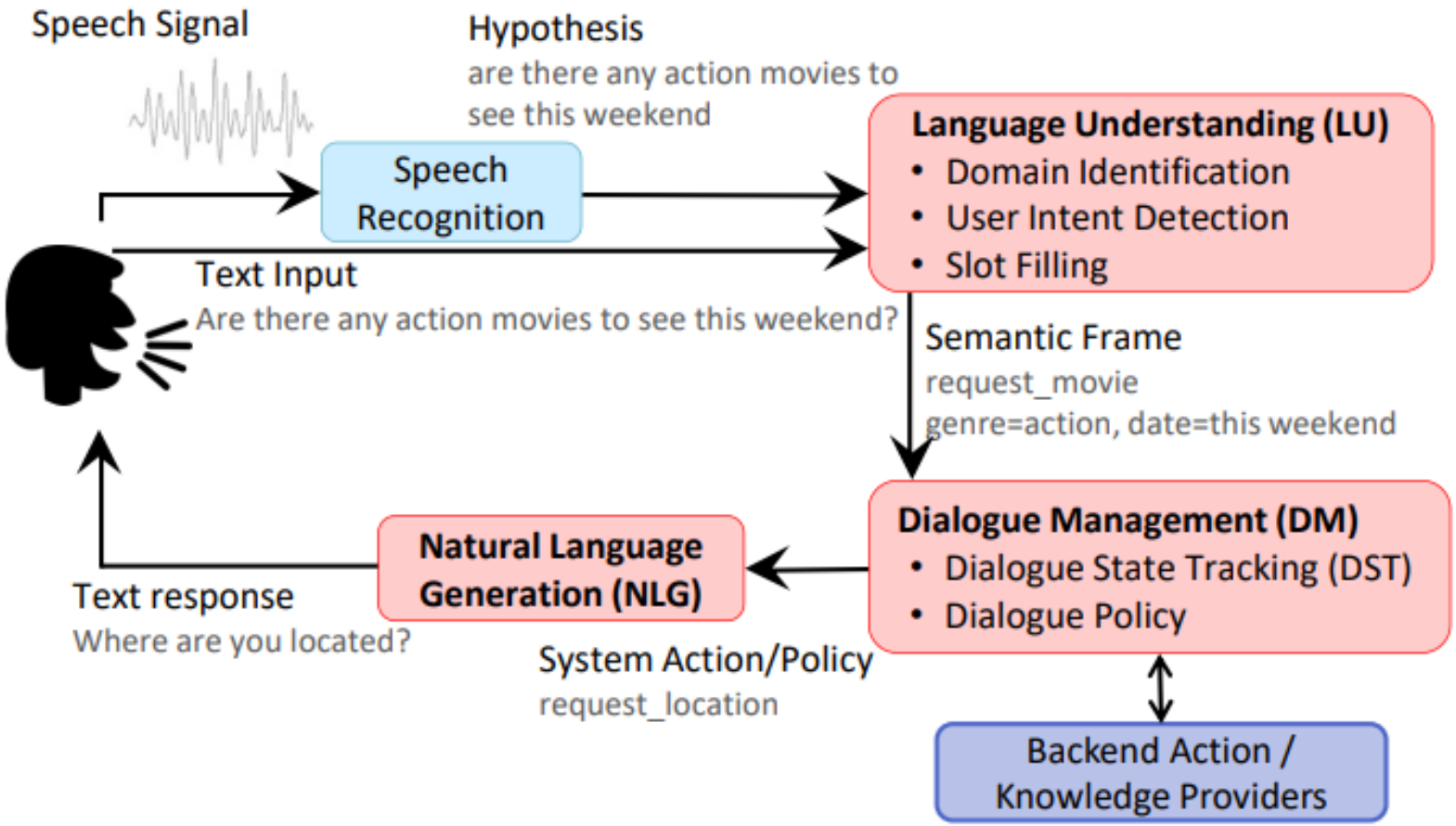
Computer Vision







Intelligent Dialog Systems

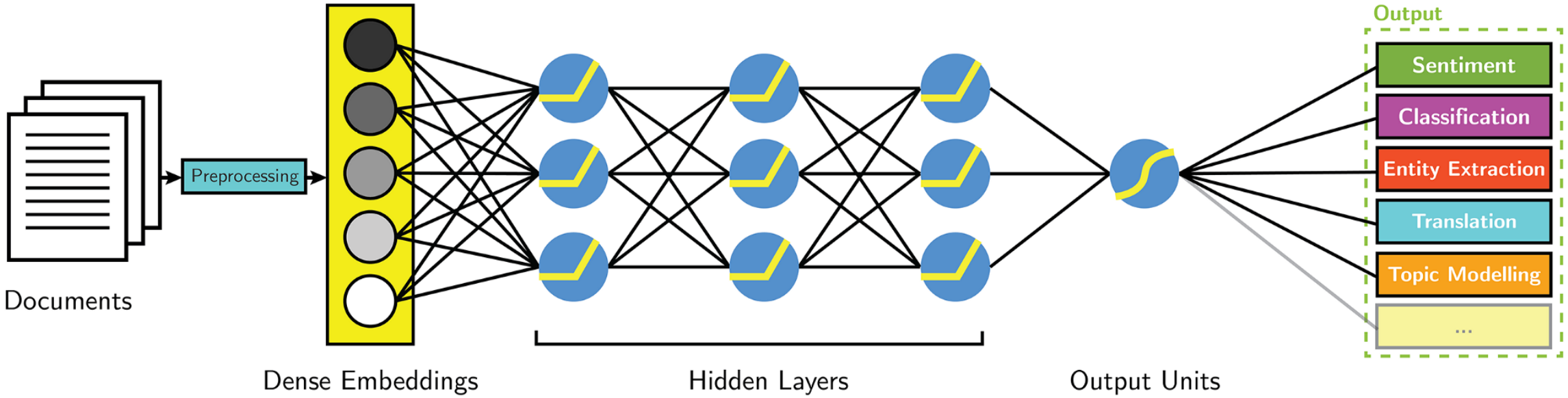


Task-oriented systems

User: I want two tickets for movie 'Deadpool'.
Agent: What is the starttime you prefer?
User: 9:00 PM
Agent: Which day?
User: Tomorrow
Agent: Which city?
User: Seattle
Agent: Which theater do you prefer?
User: AMC Pacific Place 11 theater.
Agent: Okay, I will issue 2 tickets for you, tomorrow 9:00 pm at AMC pacific place 11 theater, Seattle, movie 'Deadpool'.
User: Okay, thanks.
Agent: Thanks!

User: Do you any ticket for movie 'Hail Caesar' at theater AMC Lowes Oak Tree 6?
Agent: What start time do you prefer?
User: **7:15 pm** Tomorrow
Agent: Which city you are looking for?
User: Seattle
Agent: Okay, I will issue you tickets for movie 'Hail Caeser' tomorrow **7:50 pm** at theater AMC Lowes Oak Tree 6, Seattle.
User: Oh, no.
Agent: **It is 7:50 pm, right?**
User: Two tickets for 7:15 pm, please.
Agent: Okay, I will issue you two tickets for movie 'Hail Caeser' tomorrow **7:15 pm** at theater AMC Lowes Oak Tree 6, Seattle.
User: Thanks!
Agent: Thanks!

Natural Language Processing

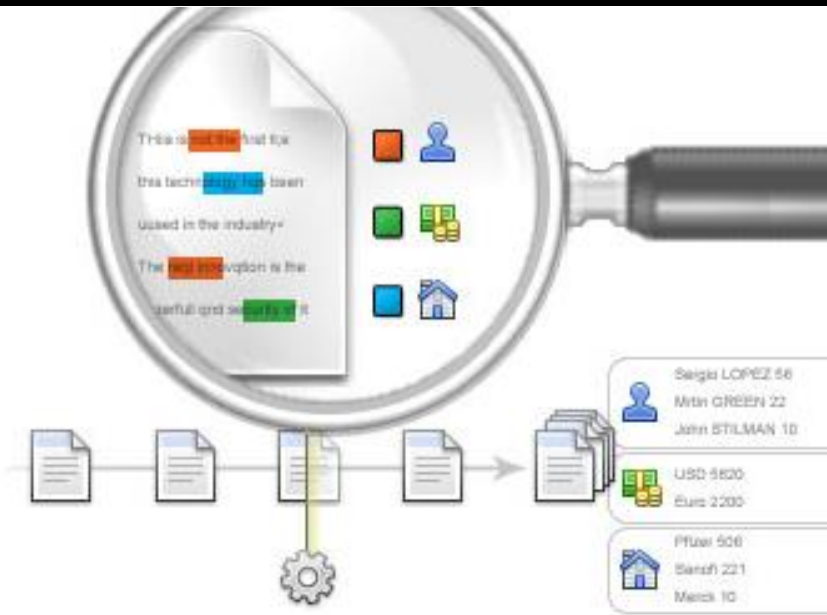


Intent Classification

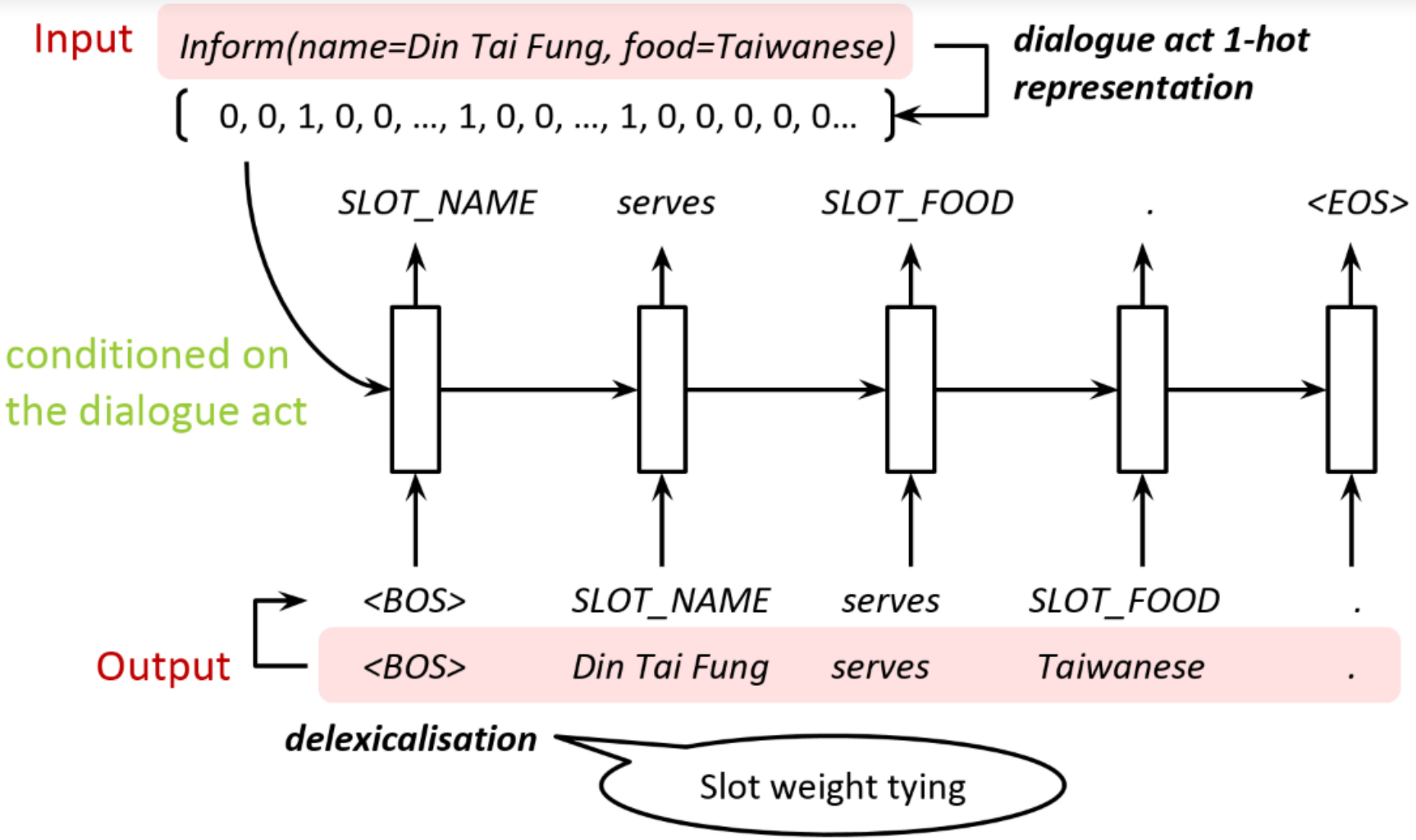
- Хочу заказать VIP такси по адресу Алтуфьевское шоссе 51 → taxi_order
- Оформите бронь с 5 по 20 августа в отеле Marriot на имя Андрея Павловича → hotel_booking
- Доставьте мне, пожалуйста, две пепперони по адресу Буйко 24 → pizza_order

Named Entity Recognition

- Хочу заказать VIP такси по адресу Алтуфьевское шоссе 51
- Оформите бронь с 5 по 20 августа в отеле Marriot на имя Андрея Павловича
- Доставьте мне, пожалуйста, две пепперони по адресу Буйко 24



Natural Language Generation



Где изучить?

coursera

UDACITY

kaggle

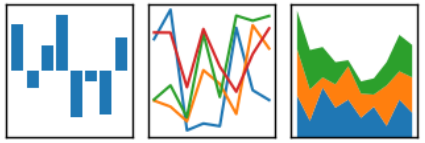
Google

Machine Learning Crash Course

Что использовать?

pandas

$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



matplotlib

Спасибо за внимание
Q&A

Овчинников Илья

ilya.ovchinnikov.ds@gmail.com

DASHA | AI