Stanford University

CS 329P : Practical Machine Learning (2021 Fall)

3.1 ML Model Overview

Qingqing Huang, Mu Li, Alex Smola

https://c.d2l.ai/stanford-cs329p

Types of ML Algorithms

Supervised

- Train on labeled data to predict labels

Semi-supervised

- Train on both labeled and unlabeled data, use models to infer labels for unlabeled data
  - E.g. self-training

Unsupervised

- Train on unlabeled data
  - E.g. clustering, density estimation

Reinforcement learning

- Use observations from the interaction with the environment to take actions to maximize reward

Types of ML Algorithms

Supervised 

Semi-supervised 

Unsupervised 

Reinforcement learning 

- We can design supervised training tasks for unlabeled data
  - Self-supervised learning: generate labels from data. E.g. word2vec, BERT
  - GAN: generating fake data with trivial label from unlabeled data,
- Training tasks can be different from how the model is evaluated / used.

Components in Supervised Training

Model

- A parameterized function to map inputs to label
  - Model parameters VS hyper parameters
  - E.g. listing house → sale price

Loss

- The measure of how good the model does in terms of predicting the outcome
  - E.g. classification / regression / contrastive / triplet / ranking
  - E.g. $(\text{predict\_price} - \text{sale\_price})^2$

Objective

- The goal to optimize model params for
  - E.g. minimize the sum of losses over examples

Optimization

- The algorithm for solving the objective

Types of Supervised Models

Decision trees 

- Use trees to make decisions

Linear methods 

- Decision is made from a linear combination of input features

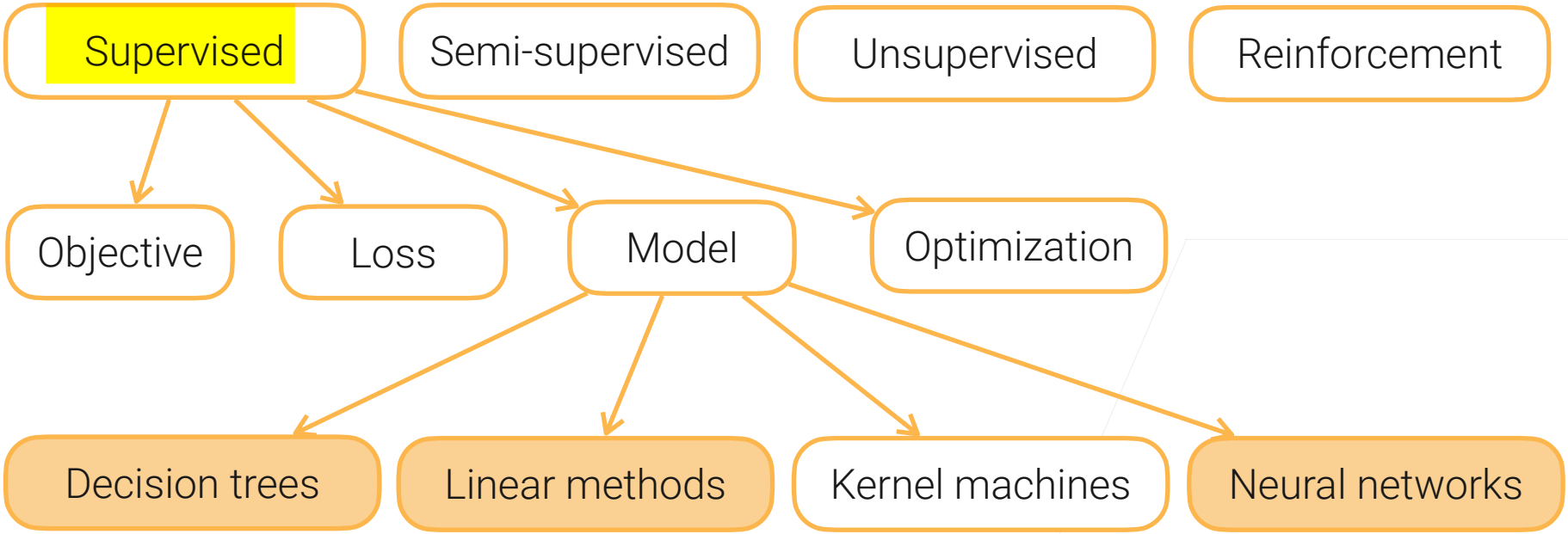
Kernel machines 

- Use kernel functions to compute feature similarities

Neural Networks 

- Use neural networks to learn feature representations

Summary


Supervised
Semi-supervised
Unsupervised
Reinforcement
Objective
Loss
Model
Optimization
Decision trees
Linear methods
Kernel machines
Neural networks

Ye