1 Introduction

本章reading:

(1) Deep Learning; (2) Deep Learning with Python; (3) Dive into Deep Learning; 的 Chapter 1.

1.1 AI, ML, DI 及历史

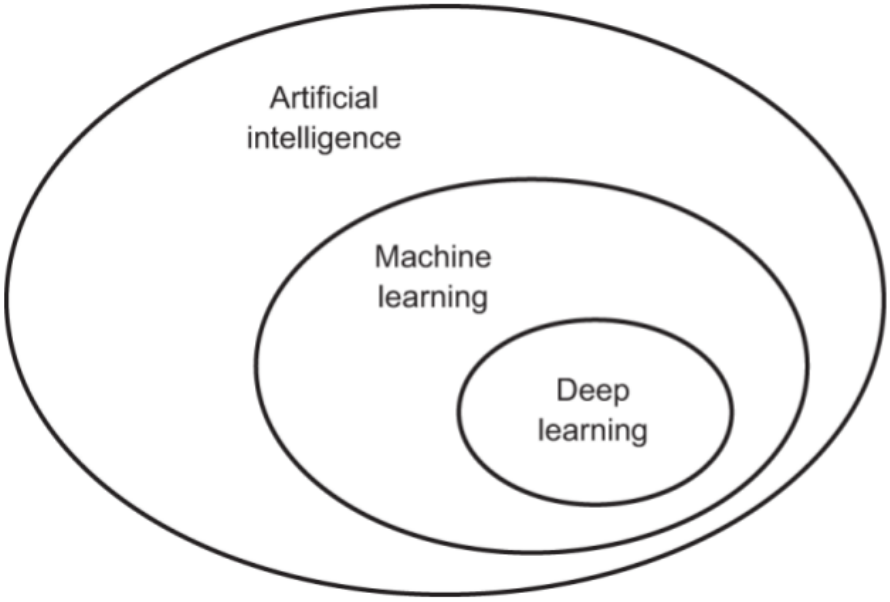


Figure 1.1 Artificial intelligence, machine learning, and deep learning

1.1.1 AI

AI 做的事情是 automate intellectual tasks normally performed by humans. ML 则是 AI 的一个分支，而 DL 又是 ML的一个分支.

Modern AI 开始于 1956 年，Dartmouth summer workshop proposal.

20 世纪 50 年代到 80 年代末，AI 的主流是 **symbolic AI (符号主义人工智能)**，即程序员为程序编写足够多的规则，通过规则的数量来模拟人. 这一方法的顶峰是 20 世纪 80 年代的**专家系统 (expert system).**

但这种方法只对于逻辑明确的问题有用. 因而就出现了新的方法叫做 **Machine Learning**.

1.1.2 ML

一个 machine learning system 是被 trained 出而不是被 explicitly programmed 出来的.

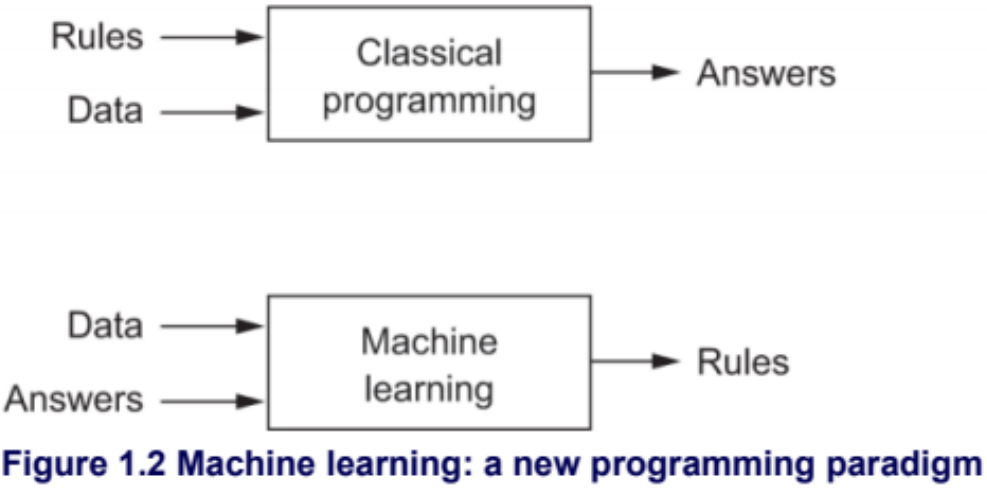


Figure 1.2 Machine learning: a new programming paradigm

ML 和数理统计关系很大，但是并不同. ML 用于处理复杂的、高维度的大型 dataset. 对如这种数据。经典的统计分析比如贝叶斯分析是不可能的. 而 ML 尤其是 DL， 用相对少的数学理论，以工程为导向进行处理.

我们需要以下这几个组件来进行 ML：

1. **Input Data Points (数据).** 比如 speech recognition 的 data points 为声音，image tagging 的 data points 为 picture.
2. **Model (模型)**: to transform the data.
3. **Objective Function (目标函数) 或叫 Loss Function:** 计算 algorithm 的 output 与expected output 的差距，检测 model 的有效性.
4. **Optimization Algorithm (优化算法)**: 接收到 Loss Function 的结果之后，调成参数，从而 optimize Loss Function,

1.1.2.1 Data 数据

每个 Dataset 由一个个 **sample(样本)** 组成，大多时候遵循 **i,i,d (independently and identically distributed，独立同分布).** sample 也叫 **data point**.

每个 sample 由一组 **features (特征)**，或叫 **covariates (协变量)** 组成. 机器学习模型会根据这些 features 进行预测. 在 **supervised learning (监督学习)** 问题中，要预测的是一个特殊的 feature，被称为 **label (标签)** 或 **target (目标)**.

比如处理图像数据时，每一张单独的照片即为一个 sample，它的 features 由每个像素数值的有序列表示. 比如，200 × 200彩色照片由200 × 200 × 3 = 120000个数值组成，“3”对应于每个空间位置的红、绿、蓝强度.

注意，**data 的 representation 是一个很重要的问题. 对于不同的 tasks，不同的数据格式是更好的. 选择更适合 tasks 的 representations 能使得 task 更加简单.** 比如下面这个数据集，极坐标就比笛卡尔坐标系更容易分割.

当每个 sample 的 feature 类别的数量都是相同的时候，其vector 是 fixed-length 的，这个长度被称为数据的 dimensionality (维数).

fixed-length 的 vector 是很适合学习的，但是并不是所有的数据都可以用 fixed-length 的 vector 表示. 比如来自互联网的分辨率和形状不同的图像，以及文本数据.

与传统 ML 方法相比，DL 的一个主要优势是可以处理不同长度的数据.

1.1.2.2 Model 模型

模型就是对数据的 transform. DL 模型和经典模型的区别在于 DL 模型由 Neural Networks 交错在一起，包含了多层的 transform.

1.1.2.3 Loss Function 目标函数

当任务在试图预测数值时，最常见的损失函数是 **squared error**，即预测值与实际值之差的平方.

当试图解决分类问题时，最常见的目标函数是 **error rate**，即预测与实际情况不符的样本比例.

有些 Loss Func（如squared error）很容易被优化，有些目标（如 error rate）由于 non-differentiability 或其他复杂性难以直接优化. 这种时候通常会优化 **a surrogate objective (代替目标)**.

通常，损失函数是根据模型 parameters 定义的，并取决于dataset. 在一个数据集上，我们可以通过最小化总损失来学习模型 parameters 的最佳值. 这个数据集由一些为 training 而收集的样本组成，称为 training dataset. 然而在 training 表现良好的模型，并不一定在新数据集上有同样的性能，因而我们需要 test dataset.

所以我们一般把 Dataset 分成两部分：**training dataset** 用于拟合模型参数，**test dataset** 用于评估拟合的模型. 当一个模型在 training set 上表现良好，但不能推广到 test set 时，这个模型被称为 **overfitting (过拟合)** 的.

1.1.2.4 Optimization Algorithm 优化算法

优化算法搜索出 loss func 的最佳 parameters，从而 minimizing loss func.

DL 中大部分流行的 Optim Algo 都基于 Gradient Descent approach. 在 Gradient Descent Approach 在每个步骤都会检查每个 parameter，看看对于某一个 parameter 如果仅改动它的话 loss 会朝哪个方向移动，然后在减少 loss 的方向上进行优化.

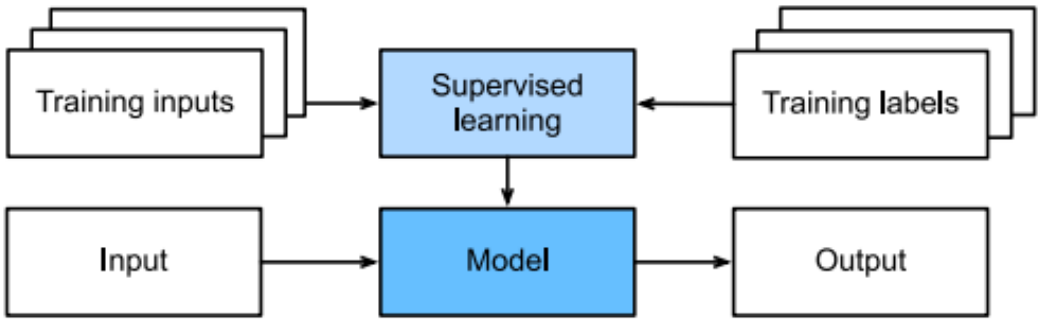
1.1.3 ML 问题类型

1.1.3.1 Supervised Learning 监督学习

supervised learning 在“给定输入 feature ”的情况下预测 labels. 每个 “feature-label” 对都称为一个 sample.

而 supervised learning 的目标是生成一个 model 能够将任何 input feature 映射到 label 上，形成预测.

比如为了预测患者的心脏病是否会发作，观察结果 "发作与否" 是 label，而患者的各项身体指标是 input features.



regression (回归)，classification (分类)， tagging (标记)， search (搜索)，recommender system (推荐系统)，sequence learning (序列学习，如机器翻译，文本到语音等) 就是典型的 supervised learning.

1.1.3.2 Unsupervised and Self-Supervised Learning 无监督学习和自监督学习

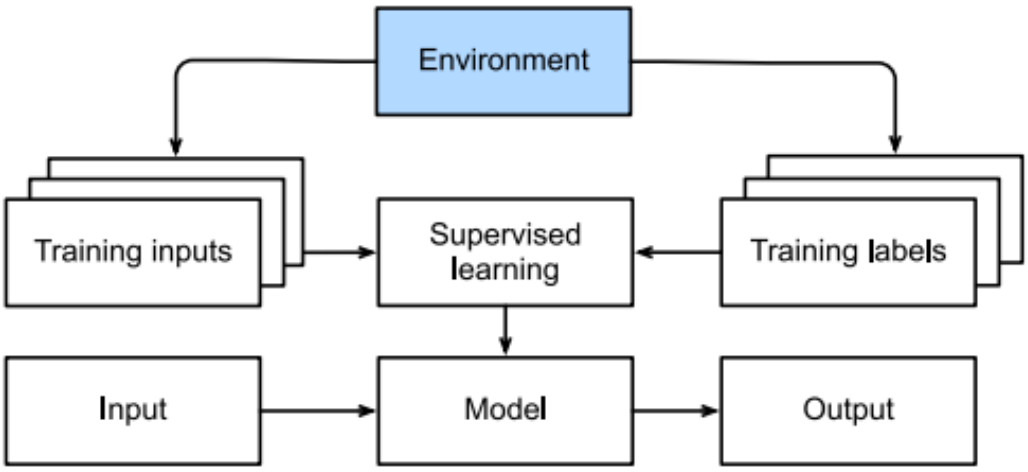
Supervised Learning 需要向模型提供巨大的数据集，每个样本包含 features 和相应的 label 值. 这些 labels 监督如何 learning.

而如果数据中没有 labels (targets)，这样的问题就叫做 unsupervised learning.

这包含了 clustering (聚类)，PCA (主成分分析)，causality (因果关系) 和 probabilistic graphical models (概率图问题) 和 GNN (generative adversarial networks, 生成对抗网络) 等问题.

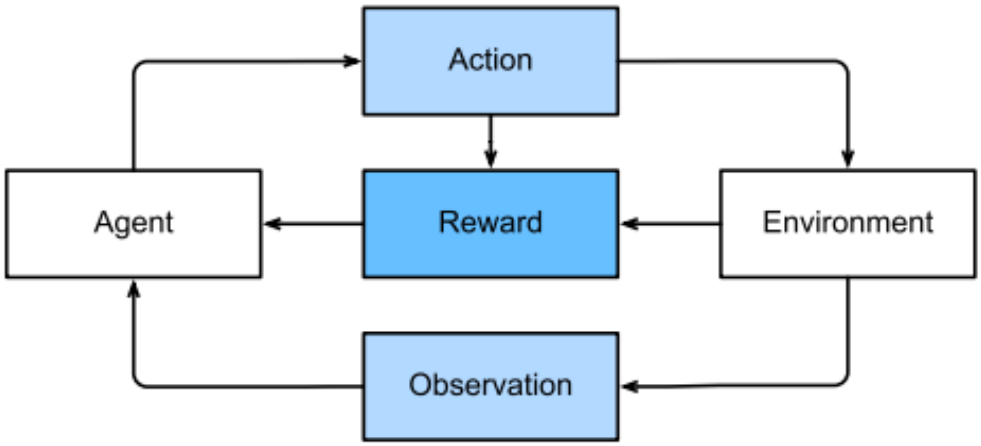
1.1.3.3 Reinforce Learning: Interacting with an Environment 强化学习: 与环境互动

不管是 supervised 还是 unsupervised learning，我们都会预先获取大量数据，然后启动模型，不再与环境交互; 这里所有 learning 都是在算法与环境断开后进行的，被称为离线学习 (offline learning). supervised learning 从环境中收集数据的模式：



而在 Reinforce Learning 问题中，agent (智能体) 在一系列的时间步骤上与环境交互. 在每个特定时间点，智能体从环境中接收一些observation，并且必须选择一个action，然后通过某种 actuator (执行器) 传输回环境，从环境中获得 reward, 开始新一轮循环. Reinforce Learning 的目标是产生一个好 policy.

我们可以将任何监督学习问题转化为强化学习问题.



当环境可被完全观察到时，强化学习问题被称为 Markov decision process (马尔科夫决策过程).

1.1.4 DL

传统的 ML 的 model 一般只进行一次对数据的 transform, 而 DL 的意思是就是多个 Layer，进行多层的深度 transformation. 因而 **DL 就是现在对于 Neural Network 的一个新的命名.**

Neural Network 的研究经历了三次浪潮：

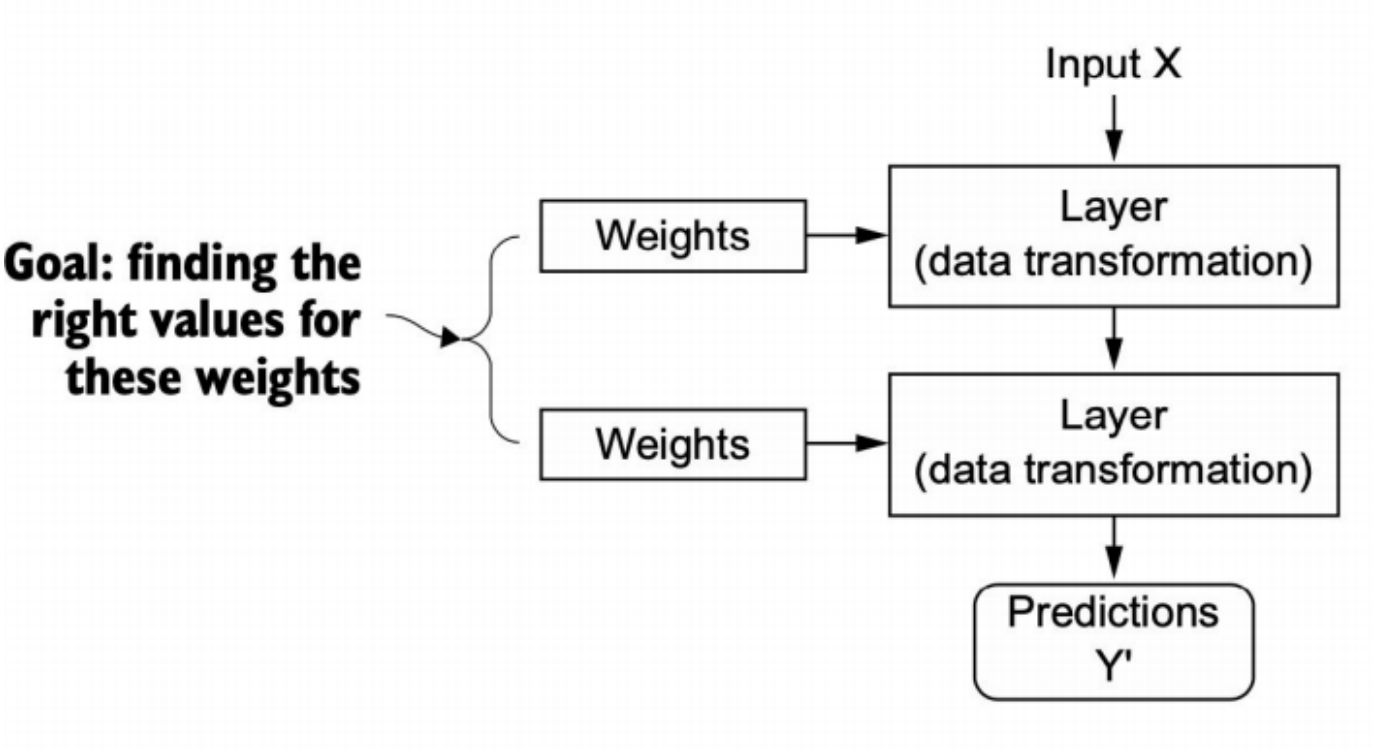
第一次是在 1940s-60s: cybernetics (控制论)

第二次在1980s-90s: connectionism (联结主义)

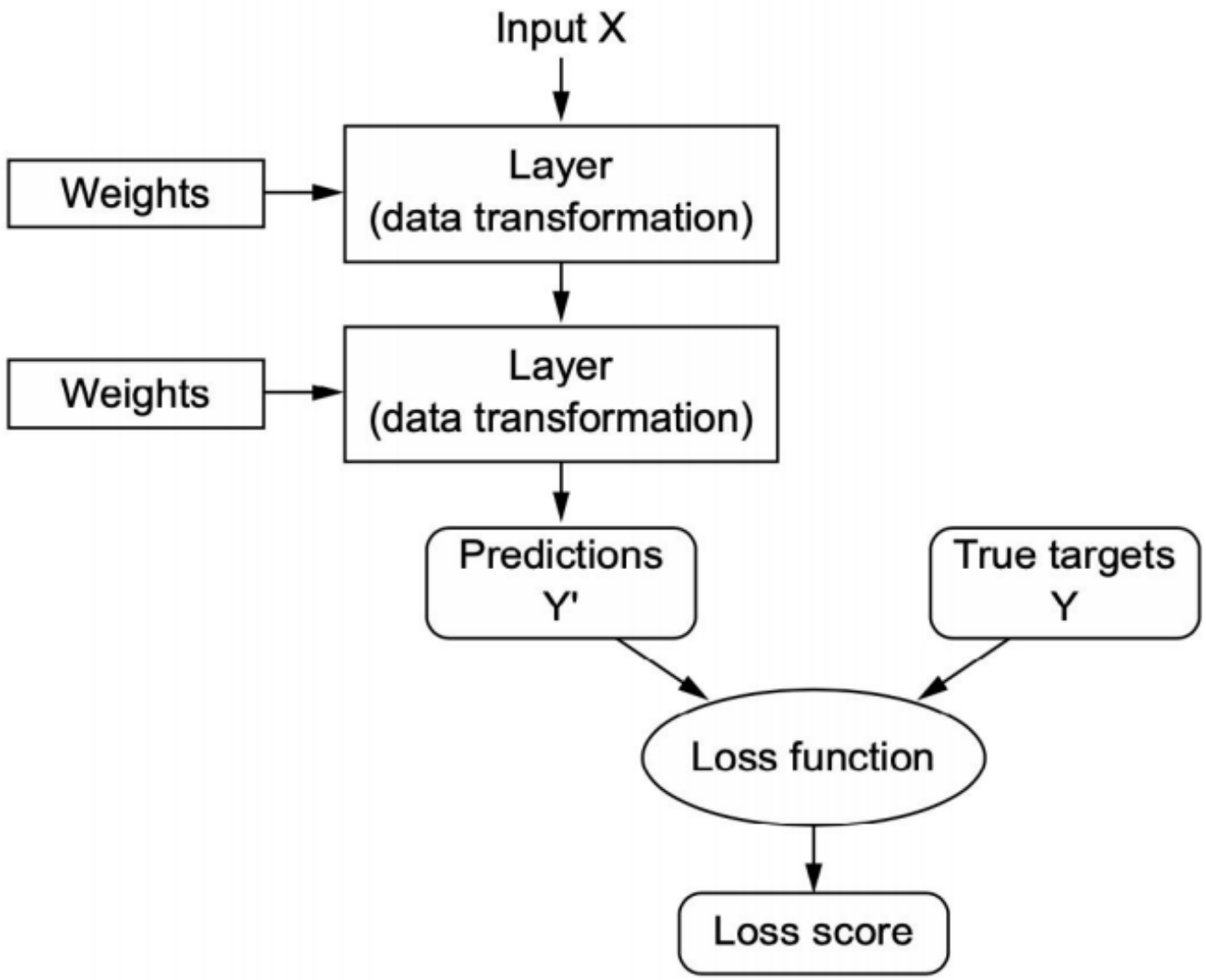
第三次是在 2006，至今: deep learning

DL 的步骤大概就是如下：

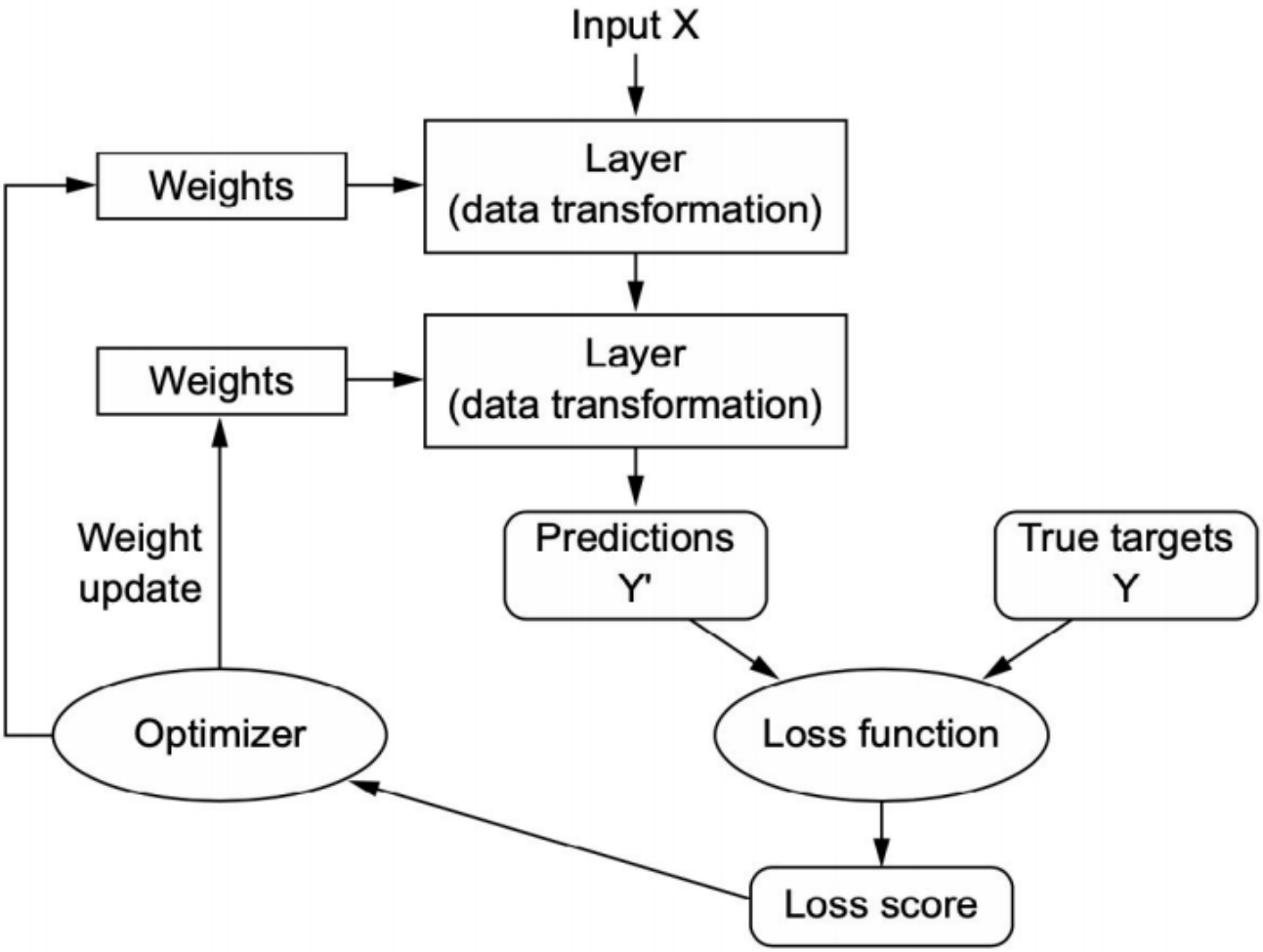
1. 根据现有的 weights (参数) 运行 model, 深度 transform



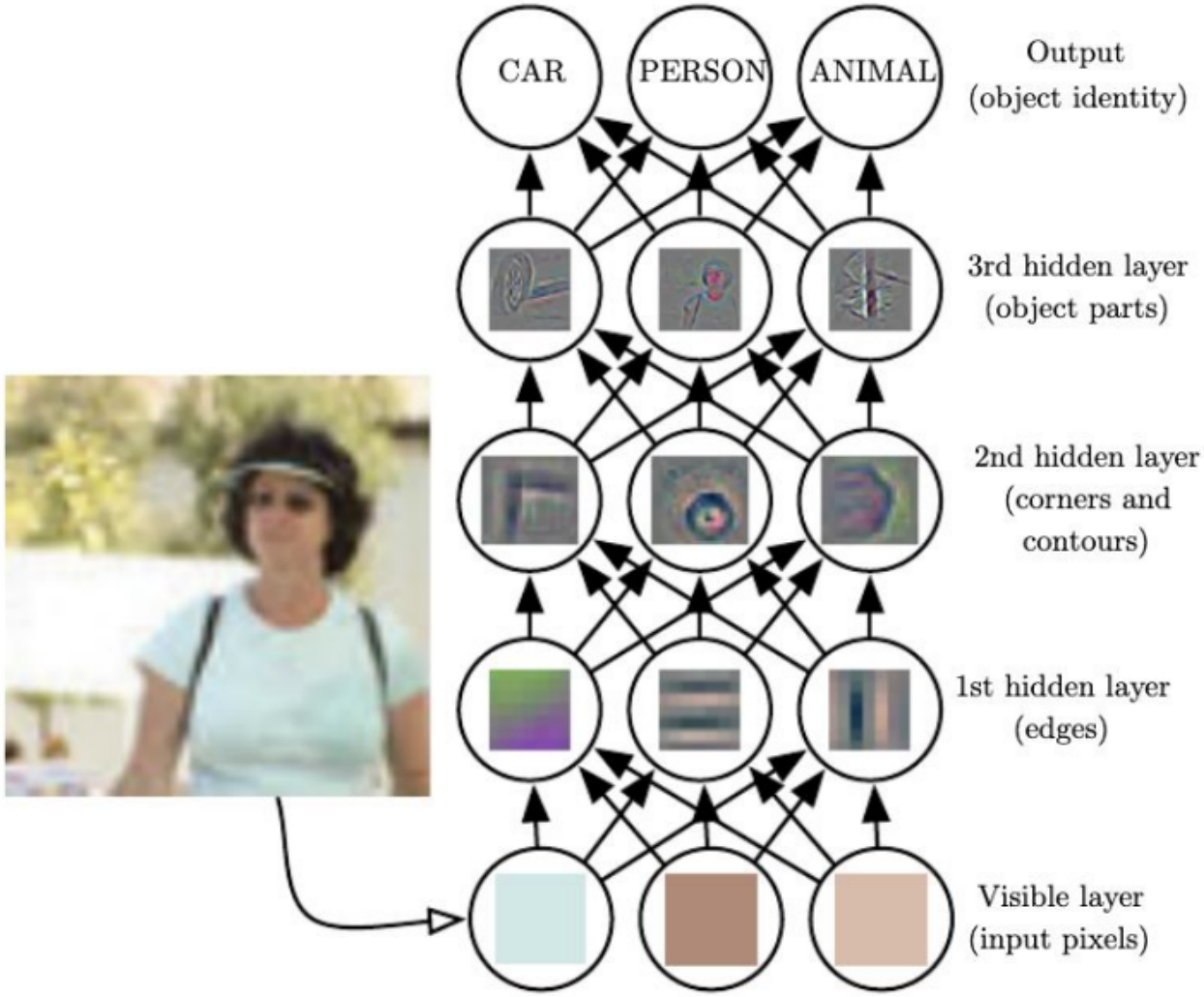
2. 将 output 的 prediction 结果传进 loss function 评估模型参数.



3. 通过 optimization algorithm 重新决定 model 的 weights.



从 input layer 到 output layer，中间的 hidden layer 总是一层一层地从抽象到具体.


CAR
PERSON
ANIMAL
Output
(object identity)
3rd hidden layer
(object parts)
2nd hidden layer
(corners and
contours)
1st hidden layer
(edges)
Visible layer
(input pixels)