什么是机器学习

机器学习现阶段指的是一种数据驱动（data-driven）的学习方式。数据驱动的学习指的是通过**给定的输入数据**和**期望获得的输出数据**来设计一个模型（模型可以被理解为是一个计算过程/推理方法）。这个模型可以通过输入的数据算出尽可能准确的输出的数据。

> 举例：通过一个人的运动量、饮食健康程度来推算一个人的寿命。这个过程中**运动量**和**饮食健康程度**就应该是输入数据，**真实的寿命**就是正确的结果，**推算出来的寿命**就是模型的输出。设计（或优化/训练）这个模型指的就是通过改变模型的某些参数，使得模型的输出尽可能接近真实值。

可以看出，要训练一个模型，必须要有可供训练的数据，这就是**数据集**。换句话说，机器学习就是从已有的数据中找到规律。

相比之下，传统的方法很多都不是数据驱动，例如信号处理中的傅立叶变换的提出就不依赖任何数据。因此，机器学习往往会比传统方法更具有“针对性”，也就是说，机器学习模型对于给定数据有更好的表现，但是传统方法具有更好的普适性。因此，当给定的训练数据数量少且不具有代表性的时候，机器学习模型就会有很多局限性。但是当训练很多且具有代表性的时候，机器学习就会超过传统的方法。随着传感器的普及、硬件算力的增加，可以被收集和处理的数据越来越多，这也就是为什么近年来机器学习逐渐占据主导。

数学语言

机器学习模型的数学描述如下：

- 已知
  - 输入数据 $\boldsymbol{x}_i \in \mathbb{R}^M,\ i = 1, 2, \dots, E$
  - 目标数据 $\boldsymbol{y}_i \in \mathbb{R}^N,\ i = 1, 2, \dots, E$
  - 机器学习模型 $\boldsymbol{f_w}(\cdot):\ \mathbb{R}^M \mapsto \mathbb{R}^N$

  其中$i = 1, \dots, E$表示一共有E个训练实例（Example），或者叫E个采样点。每一个采样点有M个特征。对于每个实例，我们想利用特征来计算N个值。计算的过程用$f_w(\cdot)$来表示，其中w就是需要被训练/设计的模型参数。

- 优化目标

  $$\underset{\boldsymbol{w}}{\text{minimize}} \quad \sum_i \mathcal{L}\{f_{\boldsymbol{w}}(\boldsymbol{x}_i) - \boldsymbol{y}_i\}$$

  其中$\mathcal{L}\{\cdot\}$表示的是损失函数（loss function），用来衡量模型输出和真实值之间的差别，最直观的loss function就是$\left|f_{\boldsymbol{w}}(\boldsymbol{x}_i) - \boldsymbol{y}_i\right|$，用差值的绝对值来表示差别。当然比如$\left(f_{\boldsymbol{w}}(\boldsymbol{x}_i) - \boldsymbol{y}_i\right)^2$也可以实现类似的功能。通常的loss function多种多样，有时也可以自己设计。但是最终的目标是让模型的输出尽可能等于真实值。

举例：我们想通过一个人的运动量x_1，饮食健康程度x_2，以及睡眠时长x_3来估计一个人的血糖y_1和血压y_2。为了实现这个目的，我们暂时用最简单的加权乘法来计算，也就是

$$\begin{aligned}\hat{y}_1 &= w_{11}x_1 + w_{12}x_2 + w_{13}x_3\\ \hat{y}_2 &= w_{21}x_1 + w_{22}x_2 + w_{23}x_3\end{aligned}$$

可以简化成

$$\underbrace{\begin{bmatrix}\hat{y}_1\\ \hat{y}_2\end{bmatrix}}_{\hat{\boldsymbol{y}}} = \underbrace{\begin{bmatrix}w_{11} & w_{12} & w_{12}\\ w_{21} & w_{22} & w_{23}\end{bmatrix}}_{W} \cdot \underbrace{\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}}_{\boldsymbol{x}}$$

简写为$\hat{\boldsymbol{y}} = W \cdot \boldsymbol{x}$。这里$\hat{(\cdot)}$表示通过模型的估计值，也就是说$\hat{\boldsymbol{y}} = f_{\boldsymbol{w}}(\boldsymbol{x}) = W \cdot \boldsymbol{x}$。进而，我们的目标就是找到合适的$\boldsymbol{w}$，使得$\hat{\boldsymbol{y}}$尽可能接近$\boldsymbol{y}$，也就是解决一个优化问题，这个优化过程往往通过数值方法实现，后面会具体解释。
如果我们希望这个模型f能够对每个人（或者说大多数人）都有效，那么我们需要收集若干人（E个人）的真实数据$(\boldsymbol{x_1}, \boldsymbol{y}_1), \dots, (\boldsymbol{x}_E, \boldsymbol{y}_E)$，然后优化$\boldsymbol{w}$，使得模型对每个训练数据都有好的表现。（这里的表现体现在loss function上）。
请注意区分x_i和$\boldsymbol{x}_i$。

机器学习的应用

当我们找到一个$\boldsymbol{w}$，使得我们能够准确的用$\boldsymbol{x}$来估计$\boldsymbol{y}$的时候，我们就可以用这个机器学习模型来处理问题了。在现实中，机器学习在很多领域都有应用，例如

- 在机器视觉中，$\boldsymbol{x}$是往往图片信息，$\boldsymbol{y}$可以是图片的分类，例如猫、狗、人等等
- 在自然语言处理中，$\boldsymbol{x}$是往往单词序列，$\boldsymbol{y}$可以是另一种语言（翻译），一种语气等等

除此之外，机器学习还有很多广阔的应用。一般来说$\boldsymbol{x}$是容易采集的信息，而$\boldsymbol{y}$是不容易测量的量。只要能训练一个模型可以从$\boldsymbol{x}$换算成$\boldsymbol{y}$，那么人们就可以很容易的估计出本来很难测量的值了。