# Ai

模型分流只需要5ms 需要nlu 200ms

评测平台 标注平台

## 蒸馏

模型蒸馏是一种机器学习技术，旨在通过将一个较大、复杂的模型（称为教师模型）的知识转移到一个较小、简化的模型（称为学生模型）上，来提高学生模型的性能。蒸馏的过程涉及到将教师模型的输出（通常是软标签，即概率分布）用于训练学生模型，以便学生模型可以更好地学习到教师模型的知识和决策能力。通过模型蒸馏，学生模型可以在保持相对较小规模的同时，具备与教师模型相近甚至更好的性能。模型蒸馏被广泛应用于许多领域，如计算机视觉、自然语言处理等

模型 蒸馏 3层 单卡300 qps 5ms

## 泛化

模型的泛化能力是指模型在未见过的数据上的表现能力。一个好的模型应该能够在新的数据样本上产生准确且可靠的预测结果。泛化能力的好坏取决于模型对训练数据的学习程度以及模型的复杂度。

为了提高模型的泛化能力，可以采取以下方法：

数据集划分：将数据集划分为训练集、验证集和测试集，用于模型的训练、调优和评估。这样可以保证模型在未见过的数据上进行验证和测试。

数据增强：通过对训练数据进行一系列的变换操作，例如旋转、翻转、缩放等，扩大训练数据集的多样性，有助于模型学习更全面的特征。

特征选择和降维：选择最具代表性的特征，去除冗余和噪声特征，可以减少模型的复杂度，提高泛化能力。

正则化技术：如L1正则化、L2正则化等，通过加入惩罚项限制模型参数的大小，防止过拟合，提高模型的泛化能力。

模型集成：通过组合多个模型的预测结果，例如投票、平均等方式，可以减少模型的方差，提高泛化能力。

总之，提高模型的泛化能力是一个综合性的问题，需要在数据准备、特征工程、模型选择和调优等多个方面进行考虑和改进。

## CUDA

异构计算

CUDA是一种异构计算的编程模型, 所谓异构计算, 就是将一个任务分开几份, 分别在不同的设备上执行. 而在CUDA编程模型中, 我们是将主要计算的部分交给GPU来完成, 而逻辑控制和数据预处理等交给CPU来完成.

专业显卡：Telsa V100 (16或32GB显存 5120个核心)

## 基本概念

NumPy(Numerical Python) 是 Python 语言的一个扩展程序库，支持大量的维度数组与矩阵运算，此外也针对数组运算提供大量的数学函数库



人工智能是最宽泛的概念,机器学习是当前一个比较有效的实现人工智能的方式,深度学习是机器学习算法中最热门的一个分支,替代了大多数的机器学习算法.

### 机器学习

归纳 演绎 训练 预测

假设空间

机器学习的本质是 参数估计

拟合一个大公式



三要素

假设空间:模型的假设或表示

优化目标:评价或损失函数

寻解算法:优化求解算法

### 深度学习

与机器学习相比,深度学习模型的复杂度难以想象.

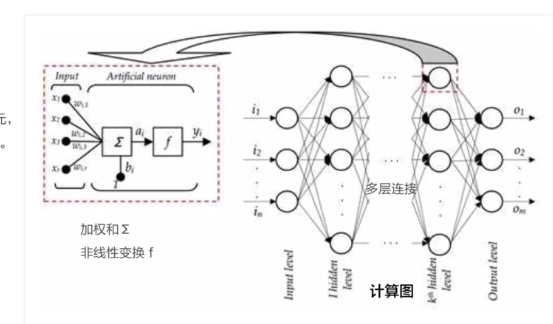
非常复杂的计算通过神经网络来完成.

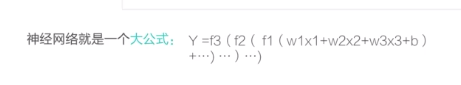
神经元:神经网络中的每个节点叫做神经元,由加权和,非线性变换(激活函数组成)

多层链接:大量的这样的节点按照不通的层次排布,形成多层的结构链接起来,即称为神经网络

前向计算:得到输出

后向传播:更新参数





因为神经网络层次很深,所以就叫深度学习咯.

本质上就是一个大公式,只不过用这种网络模型便于理解和维护.

深度学习是一个端到端的学习?不需要再去做特征工程?

也就是说不用在对这个领域由很深的理解?

深度学习的框架之下,即使你不是这个领域的专家,你弄的模型也可以像最顶尖的专家一样的水平?

## 房价预测

13个x预测一个y.

多元线性回归

屏幕快照 2021-03-07 下午4.15.25

损失函数要越小越好

五个步骤

数据处理

模型设计

训练配置

训练过程

模型保存

数据集分为训练集和测试集

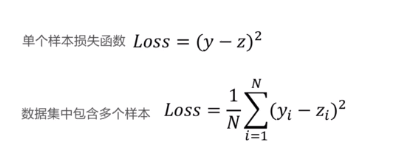
如果不拆的话, 你的模型是训练集训练出来的他能不准么...

所以测试集必须要有并且不能参与到训练中

样本归一化,就是计算出每一种数据的最大值,最小值,平均值

目前w的值设置的是随机的,这样是没有预测能力的,因为你还没有经过学习.

其实损失函数不会太多种,房价这个无非就是差异值的平方,平方的更能体验差异.



其实loss函数是以我们的w的为参数的一个函数...是这样的.