CS 316: Introduction to Deep Learning

Logistics and Introduction
Week 1

Dr Abdul Samad

Lecture Outline

- Idea of Machine Learning Software
- Key Components of Machine Learning
- Kinds of Machine Learning Problems
- Road to Deep Learning
- Examples of Deep Learning Problems

How to Design Software Solutions

Can we write programs ?

- Predict weather given geographic information, satellite images, and a trailing window of past weather.
- Given a question expressed in free-form text, answer it correctly.
- Given an image, can identify all the people it contains, drawing outlines around each.
- Present users with products that they are likely to enjoy but unlikely, to encounter.

Training Process in Machine Learning

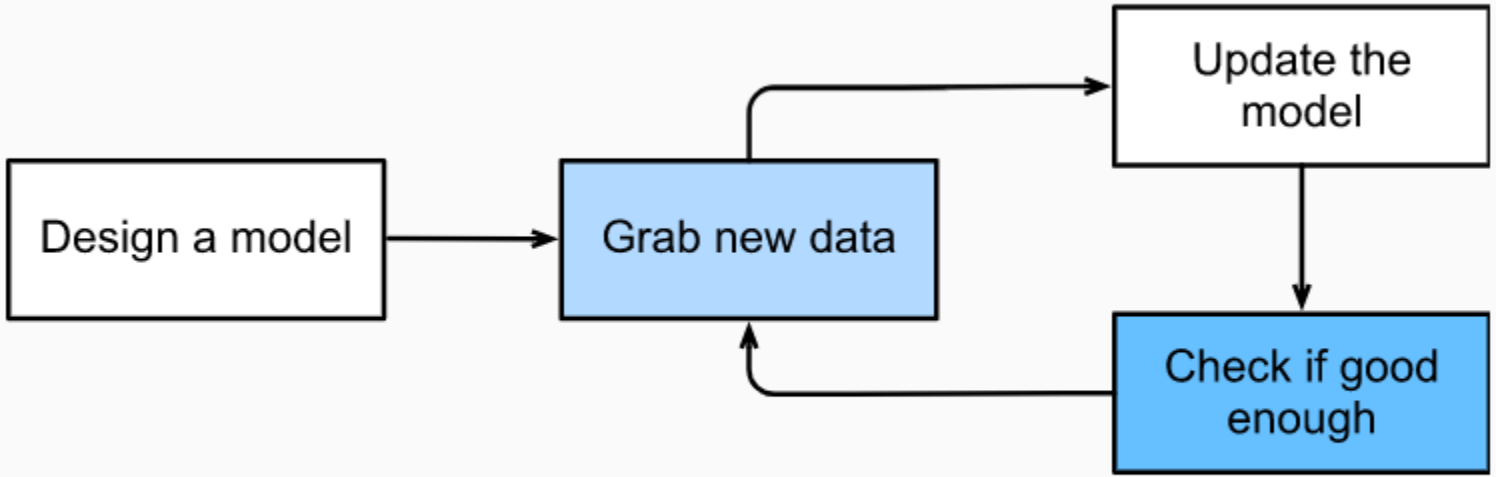
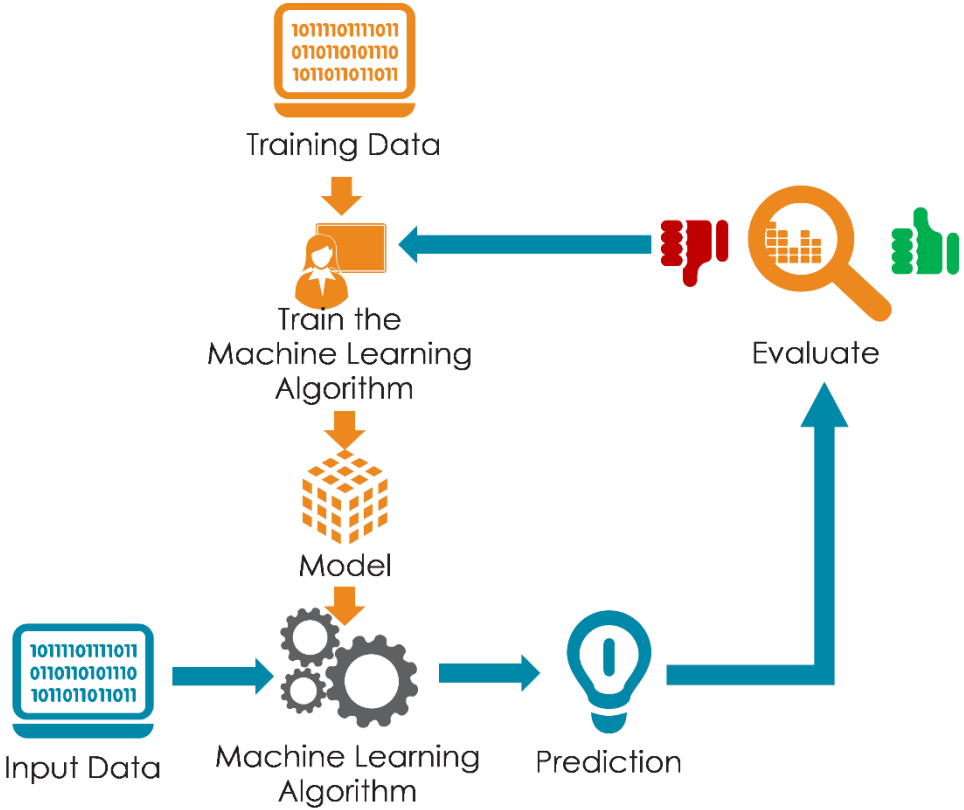


Fig. 1.1.2 A typical training process.

Key Components of Machine Learning

- Data
- Model
- Objective Function
- Optimization Algorithm



Data

- Collection of examples
- Each example consists of a set of attributes called features.
- Want to predict a special attribute that is denoted as label.
- Fixed length or variable length features.
- More data, better deep learning models.
- Garbage in , garbage out .
- Data can be biased e.g., Resume filtering

Model

- Computational machinery for ingesting data of one type, and spitting out predictions of a possibly different type.
- For example, given an image, output a label.
- Can be estimated from the data.
- Deep learning models are much complicated than traditional models.

Objective Function

- We need some formal measure for determining how good or bad our models is.
- Also denoted as Loss function.
- The lower the value of the loss function, the better the model.
- In case of regression, we have squared error.
- In case of classification, we have error rate.
- Training loss is the loss on the training data and test loss is the loss on the test dataset.
- SoftMax
- Mean Squared Error
- Cross Entropy Loss

Optimization Algorithm

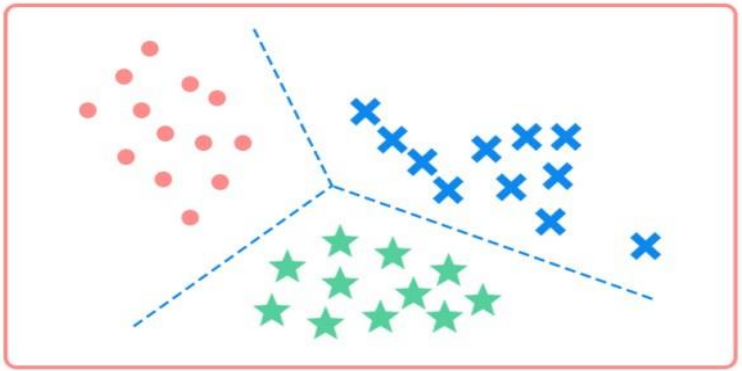
- Given a dataset, machine learning model, and a well-defined loss function, we want need an algorithm capable of searching for the best possible parameters for minimizing the loss function.
- Gradient Descent
- Adam
- AdaGrad
- Stochastic Gradient Descent
- Mini Batch Gradient Descent

Kinds of Machine Learning Problems

- Supervised Learning
  - Regression
  - Classification
- Unsupervised Learning

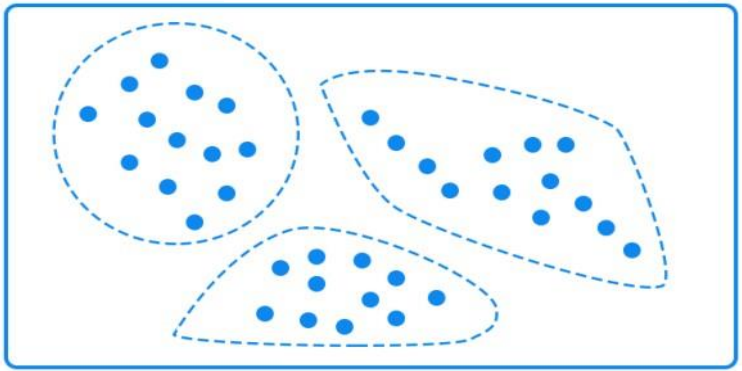
Supervised Learning vs Unsupervised Learning

Classification



Supervised Learning

Regression



Unsupervised Learning

Supervised Learning

- Predicting labels, given features.
- Each pair of (features, label) is an example.
- The goal is to produce a model to map input to labels.
- For example, given vital signs such as heart rate, blood pressure, etc. predicting Heart Attack.
- In probabilistic terms, we are estimating the conditional probability of a label given input features.

Examples of Supervised Learning

- Predict cancer vs. not cancer, given a computer tomography image.
- Predict the correct translation in French, given a sentence in English.
- Predict the price of a stock next month based on this month's financial reporting data.

Supervised Learning System Diagram

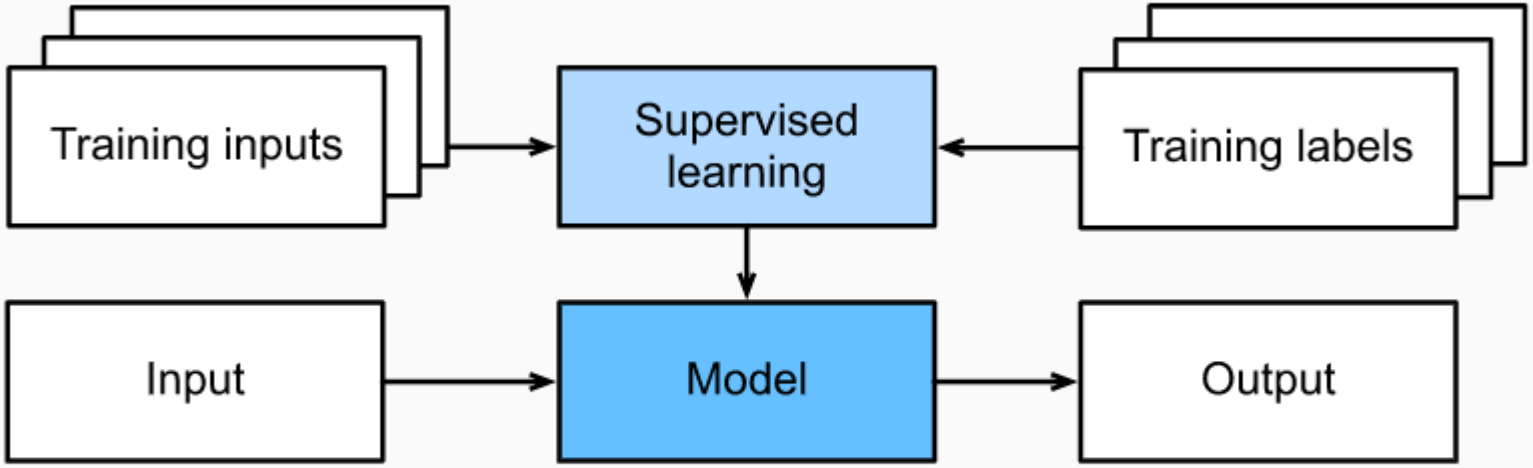


Fig. 1.3.1 Supervised learning.

Supervised Learning - Regression

- Label to predict is numerical value.
- House Price prediction
- Predict the price of a stock next month based on this month's financial reporting data
- How many hours will this surgery take?
- How much rainfall will this town have in the next six hours?

Supervised Learning - Classification

- Predict category (formally called classes)
- Binary classification {cat , dog}
- More than two classes (categories), multiclass classification problem, (cross entropy).

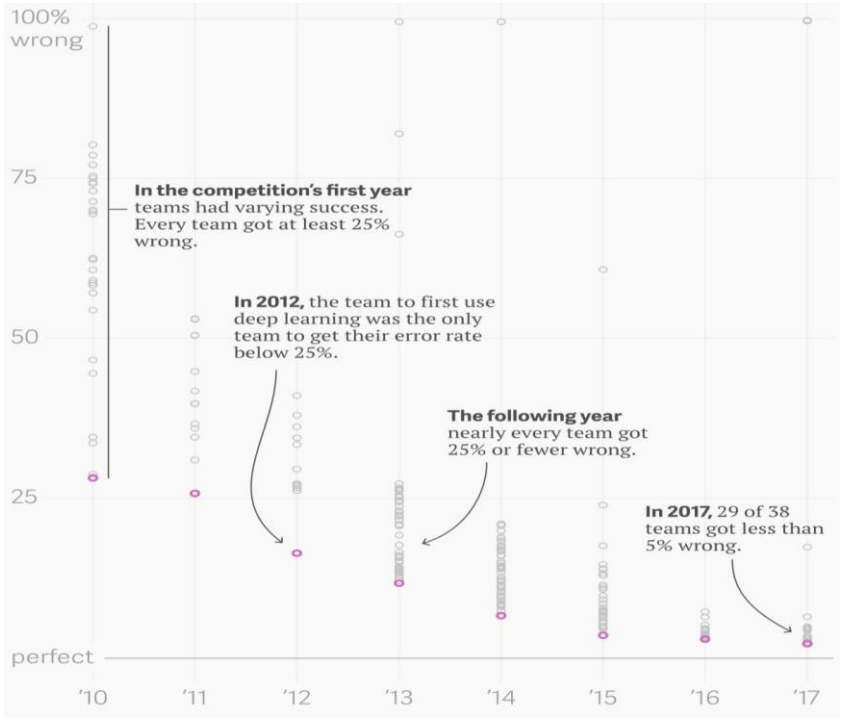
Unsupervised Learning

- No labels are given.
- Given a set of objects, cluster them into different groups based on similarity.
- Given a set of photos, can we group them into landscape photos, pictures of dogs, babies, cats, and mountain peaks.
- Given a collection of users' browsing activities, can we group them into users with similar behavior.

Road to Deep Learning

Decade	Dataset	Memory	Floating point calculations per second
1970	100 (Iris)	1 KB	100 KF (Intel 8080)
1980	1 K (House prices in Boston)	100 KB	1 MF (Intel 80186)
1990	10 K (optical character recognition)	10 MB	10 MF (Intel 80486)
2000	10 M (web pages)	100 MB	1 GF (Intel Core)
2010	10 G (advertising)	1 GB	1 TF (Nvidia C2050)
2020	1 T (social network)	100 GB	1 PF (Nvidia DGX-2)

Classifying Images



Object Detection and Segmentation

https://github.com/matterport/Mask_RCNN

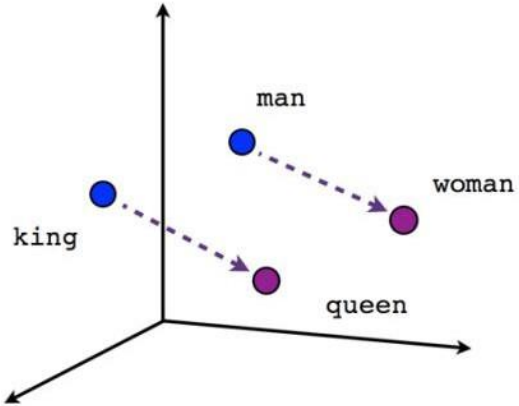
Image Style Transfer

https://github.com/zhanghang1989/MXNet-Gluon-Style-Transfer/

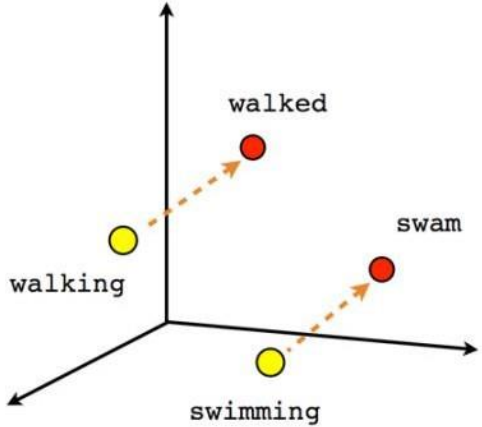
Synthesize Faces

Karras et al, ICLR 2018

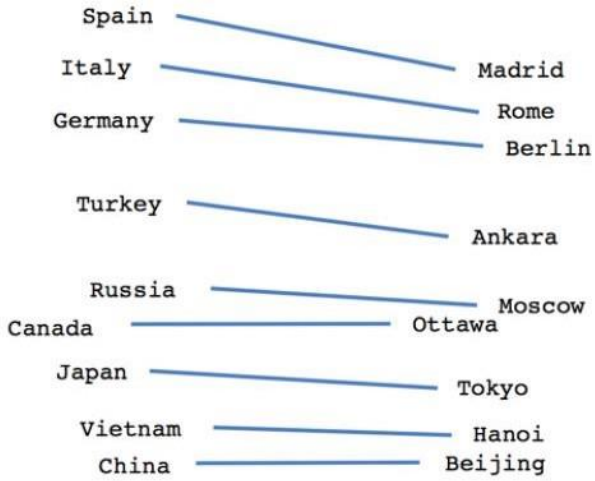
Analogies



Male-Female



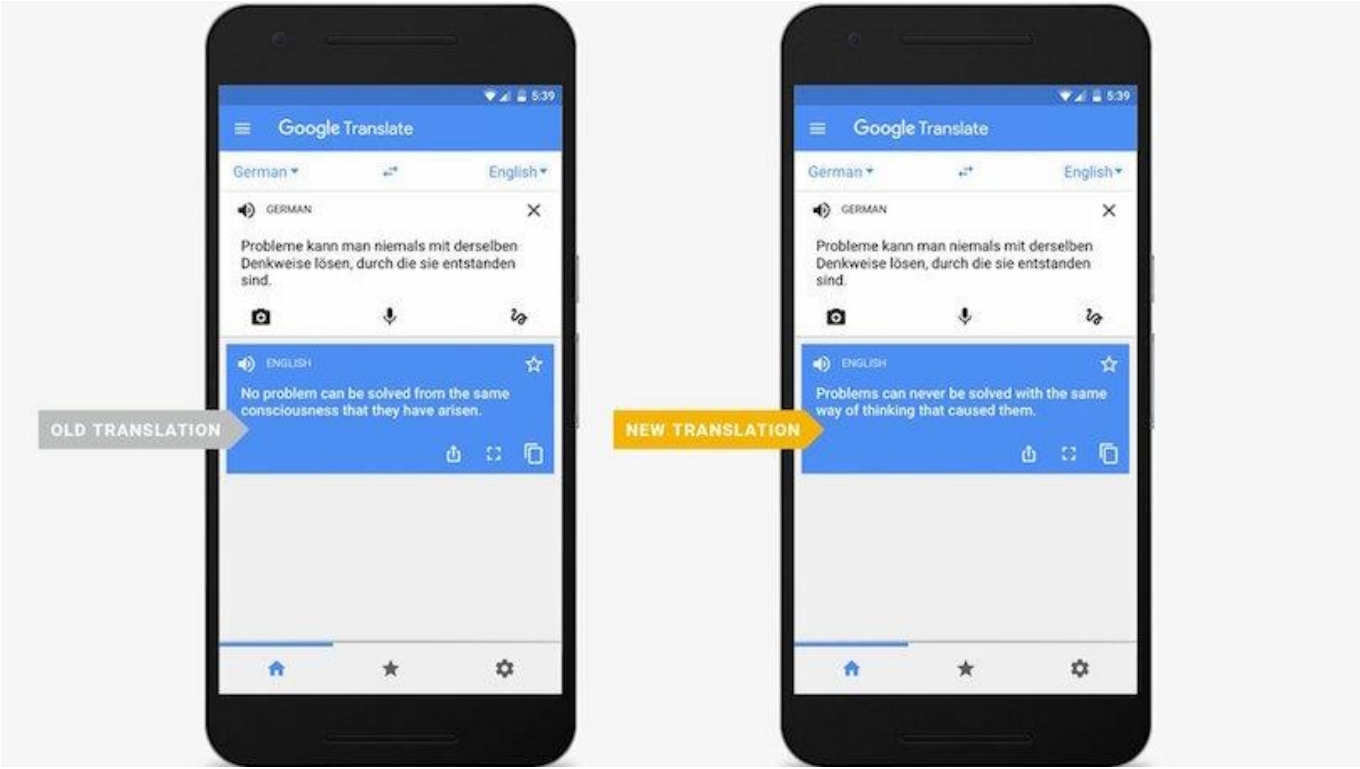
Verb tense



Country-Capital

www.tensorflow.org/tutorials/word2vec

Machine Translation

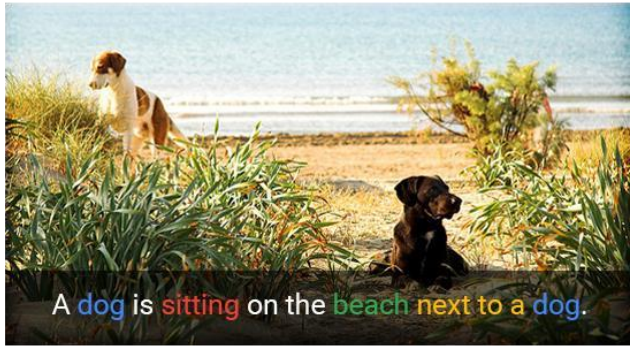


www.pcmag.com/news/349610/google-expands-neural-networks-for-language-translation

Image captioning

Human captions from the training set

Automatically captioned



Shallue et al, 2016
https://ai.googleblog.com/2016/09/show-and-tell- image-captioning-open.html

Generative Predictive Text (GPT)