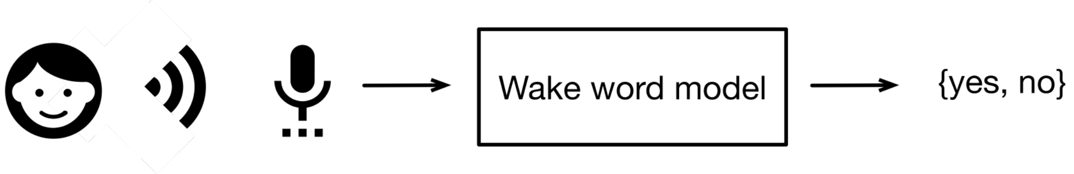
1. Introduction

Components of Application (e.g. e-commerce)

1. user interaction
2. server, infrastructure, and database
3. **business logic (brain)**

Building 'Brain' of the application

- Set of rules designed by programmers
    - it might be necessarily required when there is no accumulated experiences on the service
        - (e.g. feedback from actual users)
    - However, after enough experiences are accumulated, it might be impossible for programmers to learn all the experiences and update the software
- ML is the study of powerful techniques that can **learn from experience**
    - **As an ML algorithm accumulates more experience, its performance improves**
    - **There are certain tasks which cannot be defined by set of rules or codes**
        - **e.g. speech recognition**

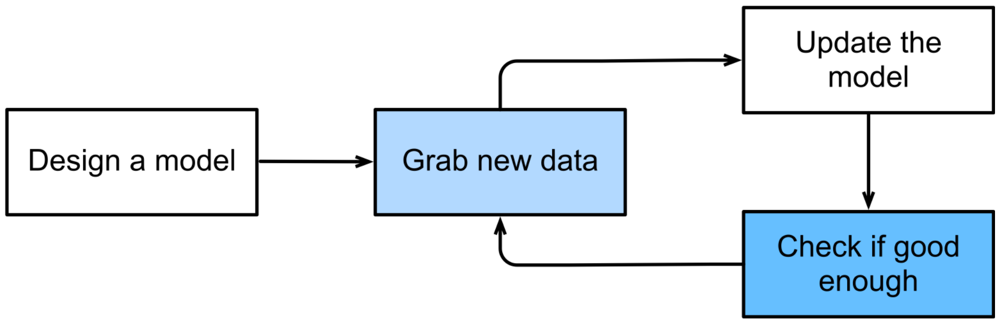


1. Introduction

Designing ML Tasks

- **Not explicitly defined program** whose decision is made by numerous **parameters**
- Using **dataset** to determine the best possible set of parameters
- After fixing the parameters, we call the program a **model.**

ML Training Process



Deep Learning Models

- Relieving labor-intensive process for feature engineering based on the domain-specific preprocessing
- DL's many-layered (or hierarchical) structure provide a capability to address low-level perceptual data in a way that shallow ML models are hard to obtain

2. The Key Components: Data, Models, and Algorithms

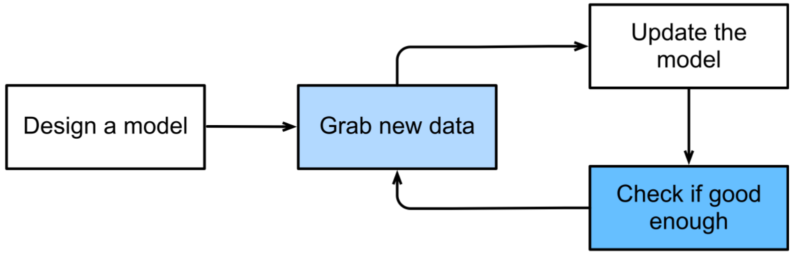
Key components

- The **data** that we can learn from.
- A **model** of how to transform the data.
- A **loss function** that quantifies the badness of our model.
- An **algorithm** to adjust the model's parameters to minimize the loss.

2.1 Data

- A collection of numerical attributes called **features**.
- In the supervised learning problems, for a special feature is designated as the prediction **target**
- Representing data into vectors
    - Data consists of fixed-length vectors and the length of the vectors is called the dimensionality of the data.
    - DL models provides methods to handle varying-length data such as sentences
- Big data
    - A major contributor to the success of modern deep learning, and many of the most exciting models in deep learning do not work without large datasets.
- Quality of data
    - One common failure mode occurs in datasets where some groups are unrepresented in the training data

2. The Key Components: Data, Models, and Algorithms



2.2 Model

- Model transforms the data into some targets
- DL models consist of many successive transformations of the data that are chained together top to bottom

2.3 Objective Function

- Formal measures of how good models are
- By convention, we usually define objective functions so that lower is better
- These functions are sometimes called loss functions or cost functions
- The output of the function is called **error**

$$f(y, \hat{y}) = (y - \hat{y})^2$$

squared error

Role of objective function

- The best values of a model's parameters are learned by minimizing the loss incurred on a training dataset

Evaluation

- Training data does not guarantee generalizability of the model
    - whether the model work well on (unseen) test data.
- So we will typically want to split the available data into two partitions:
    - **Training data** for fitting model parameters → Training error (reducing bias)
    - **Test data** which is held out for evaluation → Test error (measuring variance)

train data

test data

2. The Key Components: Data, Models, and Algorithms

2.4 Optimization algorithms

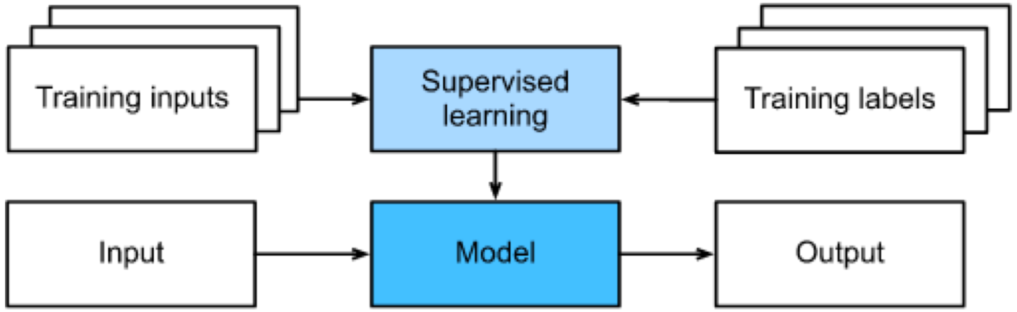
- Searching for the best possible parameters for minimizing the loss function
- Gradient descent
    - Updating parameter with just a small amount in the direction that reduces the loss.

3. Kinds of Machine Learning; Supervised learning

$\{\mathbf{x}_i, y_i\}_{i=1}^{n}$

$\mathbf{x}_i$: features of data i
y_i: label of data i
n: data instances

- Supervised learning addresses the task of predicting targets given inputs.
- Our goal is to produce a model f_θ that maps any input $\mathbf{x}_i$ to a prediction $f_\theta(\mathbf{x}_i)$ to target y_i
- In probabilistic terms, we typically are interested in maximizing the conditional probability, likelihood, $P(y|f_\theta(\mathbf{x}_i))$



- Training supervised learning models
  - Grabbing a big collection of data whose covariates (i.e. independent variables, features) are known
  - Selecting a random subset, and acquiring the ground truth labels for each subset
- Testing the model
  - Feeding unseen inputs to the learned model, using its outputs as predictions of the corresponding label

3. Kinds of Machine Learning; Supervised learning

3.1 Regression

- Predicting targets with arbitrary values in some range
  - "How many hours will this surgery take?"
  - "How many dogs are in this photo?"
- Learning models that minimize the distance between our predictions and the observed values.

L1 loss $$l(y, y') = \sum_i |y_i - y'_i|$$ noise from a Laplace distribution $$f(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\boxed{\frac{|x-\mu|}{b}}\right)$$

L2 loss $$l(y, y') = \sum_i (y_i - y'_i)^2.$$ noise from a Gaussian distribution $$f(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\boxed{\frac{(x-\mu)^2}{2\sigma^2}}\right)$$

3.2 Classification

- Predicting which category (formally called classes), among some discrete set of options
  - "Whether an object in the image is a dog or a cat?

$$p(y = \text{poisonous} \mid x) = 0.9$$

  - A classifier is 90% sure that the image depicts a cat.
  - The magnitude of the probability conveys one notion of **uncertainty**.

3. Kinds of Machine Learning; Supervised learning

3.2 Classification (Cont.)

- Loss

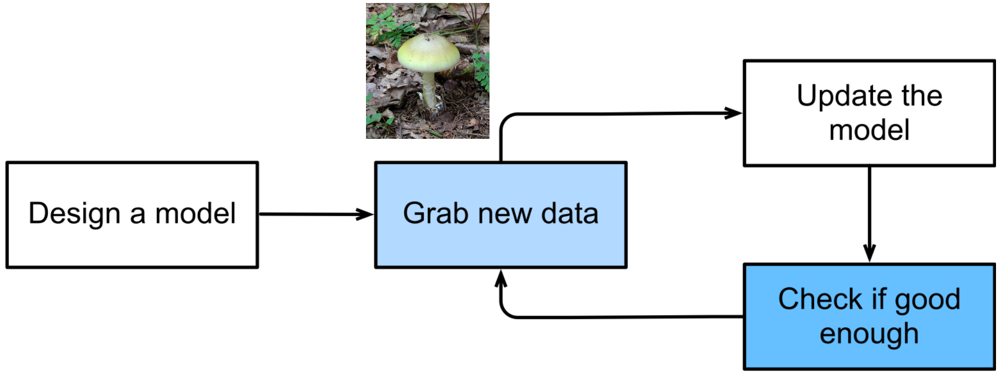
$$L(\text{action}|x) = E_{y\sim p(y|x)}[\text{loss}(\text{action}, y)].$$

$$L(a = \text{eat} \mid x, y = \text{poisonous}) = 0.9 * \infty + 0.1 * 0 = \infty$$

$$L(a = \text{discard} \mid x, y = \text{poisonous}) = 0.9 * 0 + 0.1 * 1 = 0.1$$

loss(eat, poisonous) **loss(discard, poisonous)**

- Further consideration on classification tasks
  - Multiclass classification
  - Multilabel classification
  - Hierarchical classification



poisonous normal

a = [0.9, 0.1]

y = [1, 0]

3.3 Others

- Tagging (Multilabel classification)
  - Predicting classes that are not mutually exclusive is called multilabel classification
  - Autotagging (automatically tagging keywords on a post)
  - Object detection

"machine learning", "technology", "gadgets", "programming languages", "linux", "cloud computing", "AWS".

3. Kinds of Machine Learning; Supervised learning

3.3 Others (Cont.)

- Search and ranking
  - Producing ordered subsets of elements from a larger set which are most relevant to a query
  - i.e. estimating some target score y_i, given a query q_i and document d_j.
- Recommender systems
  - Retrieving relevant items personalized to a specific user
  - i.e. estimating some target score y_i, such as the probability of purchase, given a user u_i and product p_j.
  - Problems: censoring and feedback loops
    - Relieving feedback imbalance from a personal tendency of user rating
- Sequence Learning
  - successive inputs and successive outputs.
  - Machine translation
  - POS Tagging and Parsing
  - Speech recognition