Fundamentals of Machine Learning

OVERVIEW of MACHINE LEARNING

Amit K Roy-Chowdhury

Acknowledgments: Adapted from slides at https://probml.github.io/pml-book/teaching1.html by Prof. Saw Shier Nee

What is Machine Learning?

A popular definition of **machine learning** or **ML**, due to Tom Mitchell [Mit97], is as follows:

> A computer program is said to learn from experience E with respect to some class of tasks T, and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

> Almost all of machine learning can be viewed in probabilistic terms, making probabilistic thinking fundamental. It is, of course, not the only view. But it is through this view that we can connect what we do in machine learning to every other computational science, whether that be in stochastic optimisation, control theory, operations research, econometrics, information theory, statistical physics or bio-statistics. For this reason alone, mastery of probabilistic thinking is essential.

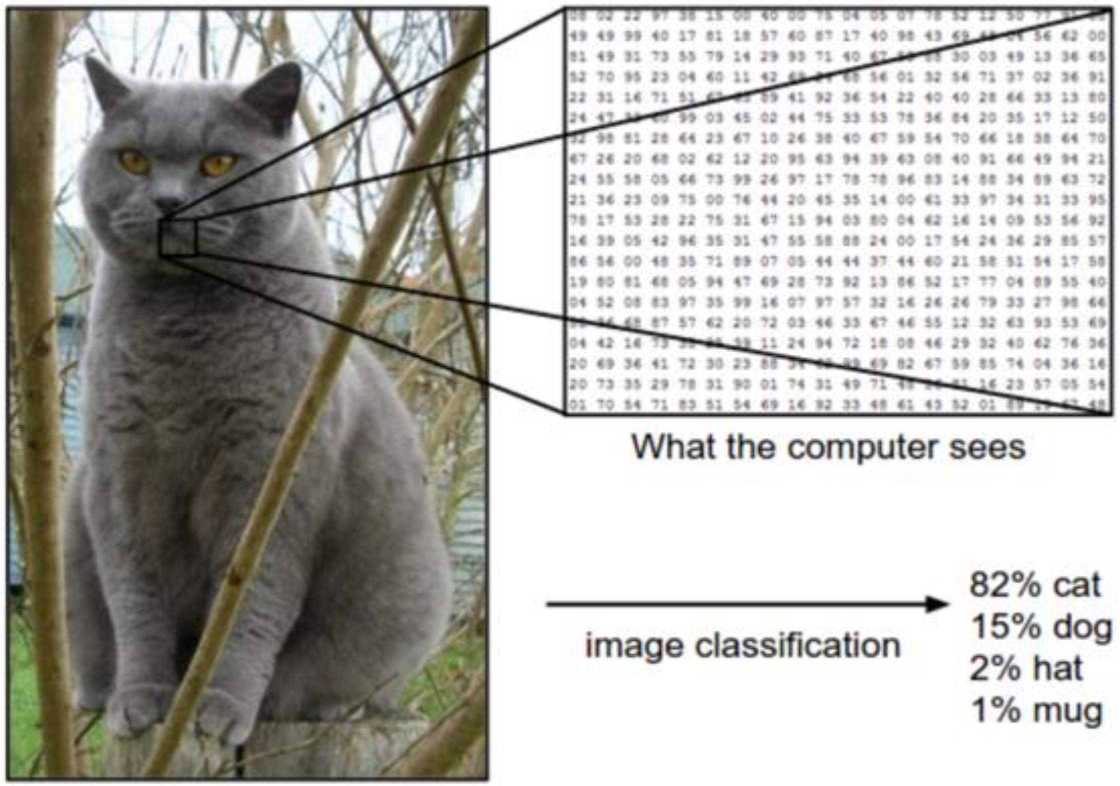
By Shakir Mohamed, research at Deep Mind.

Machine Learning Approaches

Supervised Learning	Unsupervised Learning	Reinforcement Learning
Labelled data with guidance	No labels	Interacts with environment, decide action, learns by trial and error method

Self-supervised Learning

Supervised Learning - Classification



$$input\ space, \mathcal{X} = \ set\ of\ images$$
$$outout\ space, \mathcal{Y} = \ set\ of\ classes$$

$$f: \mathcal{X} \rightarrow \mathcal{Y}$$

Supervised Learning - Classification

Empirical Risk Minimization

$$\boxed{\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \frac{1}{N}\sum_{n=1}^{N} \ell(y_n, f(\boldsymbol{x}_n; \boldsymbol{\theta}))}$$

$$\mathcal{L}(\boldsymbol{\theta}) \triangleq \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}\left(y_n \neq f(\boldsymbol{x}_n; \boldsymbol{\theta})\right)$$

$$\mathbb{I}(e) = \begin{cases} 1 & \text{if } e \text{ is true} \\ 0 & \text{if } e \text{ is false} \end{cases}$$

Supervised Learning - Classification

Uncertainty → using conditional probability distribution

$$p(y = c|\boldsymbol{x};\boldsymbol{\theta}) = S_c(f(\boldsymbol{x};\boldsymbol{\theta}))$$

Constraints:

$$0 \le f_c \le 1$$

$$\sum_{c=1}^{C} f_c = 1$$

Softmax function

$$\mathcal{S}(\boldsymbol{a}) \triangleq \left[\frac{e^{a_1}}{\sum_{c'=1}^{C} e^{a_{c'}}}, \ldots, \frac{e^{a_C}}{\sum_{c'=1}^{C} e^{a_{c'}}}\right]$$

Supervised Learning - Classification

Maximum Likelihood Estimation

Minimizing Negative Log Likelihood

$$\hat{\boldsymbol{\theta}}_{\mathrm{mle}} = \operatorname*{argmin}_{\boldsymbol{\theta}} \mathrm{NLL}(\boldsymbol{\theta})$$

$$\mathrm{NLL}(\boldsymbol{\theta}) = -\frac{1}{N}\sum_{n=1}^{N} \log p(y_n | f(\boldsymbol{x}_n; \boldsymbol{\theta}))$$

Supervised Learning - Regression

Output = Real-value → Quadratic loss

$$\ell_2(y, \hat{y}) = (y - \hat{y})^2$$

$$\mathrm{MSE}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{n=1}^{N} (y_n - f(\boldsymbol{x}_n; \boldsymbol{\theta}))^2$$

Supervised Learning - Regression

Uncertainty → Assume output distribution = Gaussian

$$\mathcal{N}(y|\mu,\sigma^2) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

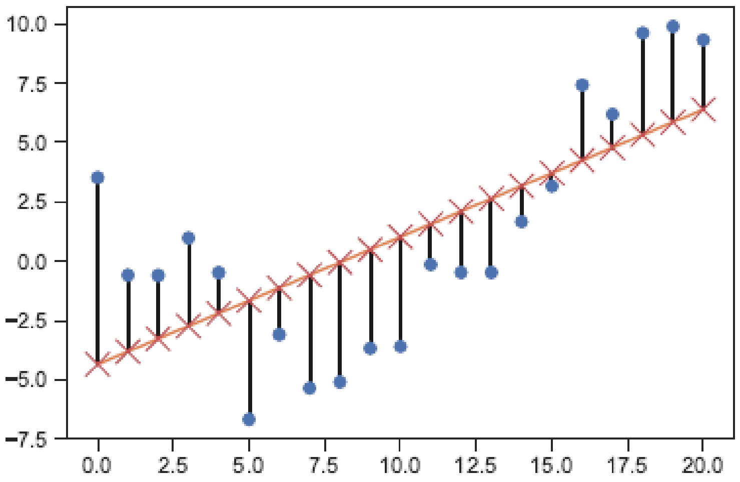
$$p(y|\boldsymbol{x};\boldsymbol{\theta}) = \mathcal{N}(y|f(\boldsymbol{x};\boldsymbol{\theta}),\sigma^2)$$

$$\begin{aligned}\mathrm{NLL}(\boldsymbol{\theta}) &= -\sum_{n=1}^{N}\log\left[\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2}(y_n - f(\boldsymbol{x}_n;\boldsymbol{\theta}))^2\right)\right] \\ &= \frac{N}{2\sigma^2}\mathrm{MSE}(\boldsymbol{\theta}) + \mathrm{const}\end{aligned}$$

Supervised Learning - Regression

Linear Regression

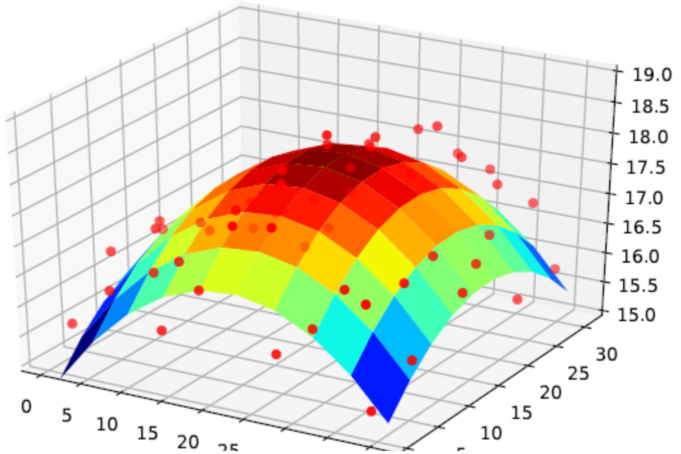
1 Feature



$$f(x;\boldsymbol{\theta}) = b + wx$$

Polynomial Regression

Features Engineering



$$f(\boldsymbol{x};\boldsymbol{w}) = w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2$$

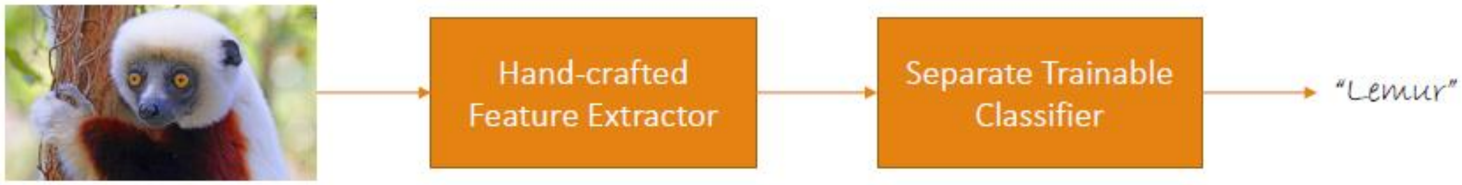
Deep Neural Network

Feature Extraction Automatically

$$f(\boldsymbol{x};\boldsymbol{w},\mathbf{V}) = \boldsymbol{w}^{\mathsf{T}}\phi(\boldsymbol{x};\mathbf{V})$$

Why Deep Learning?

- Traditional pattern recognition



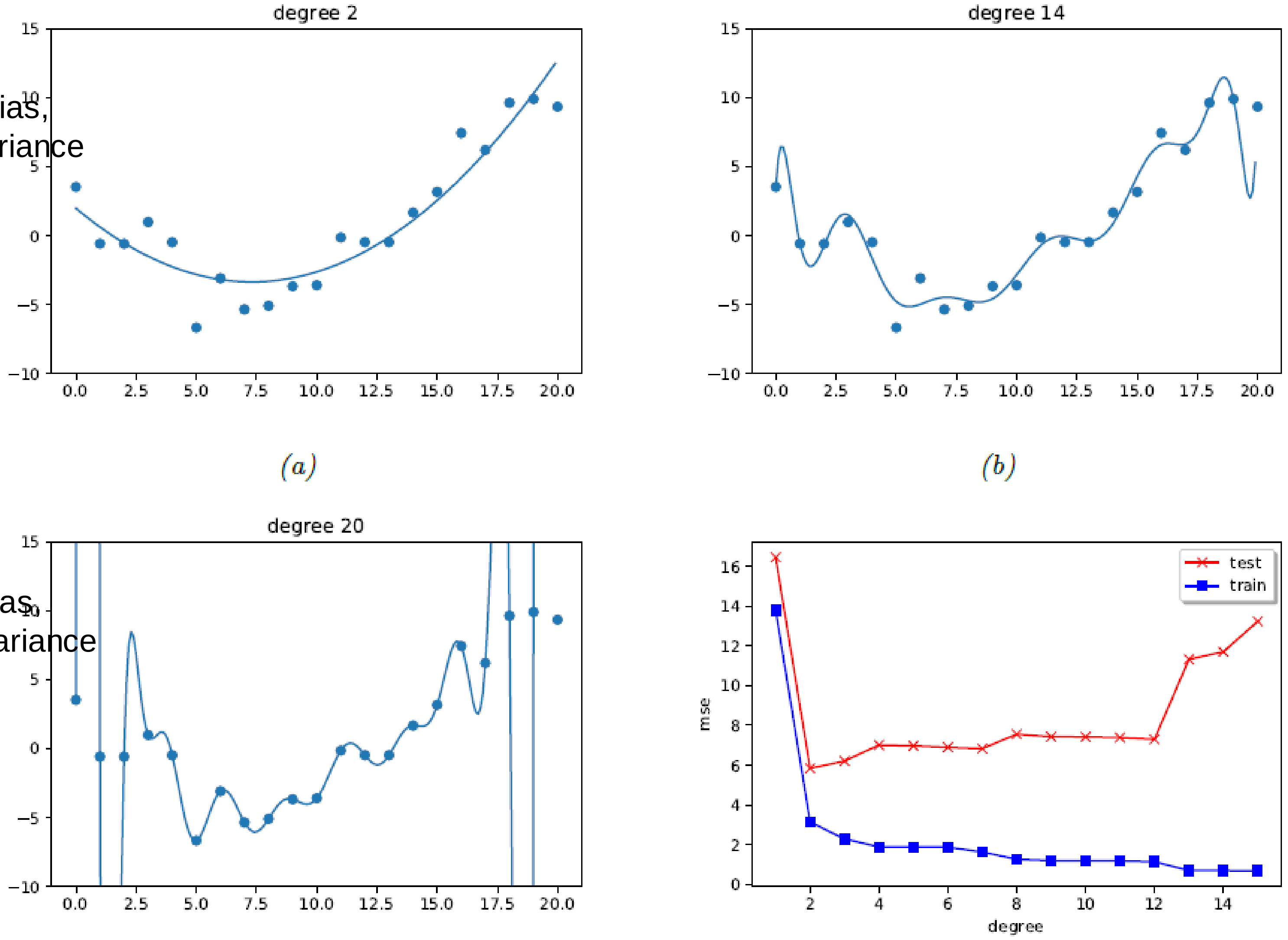
- End-to-end learning → Features are also learned from data

Slide Courtesy : Efstratios Gavves - UVA Deep Learning

Generalization

High bias,
low variance

Low bias,
high variance



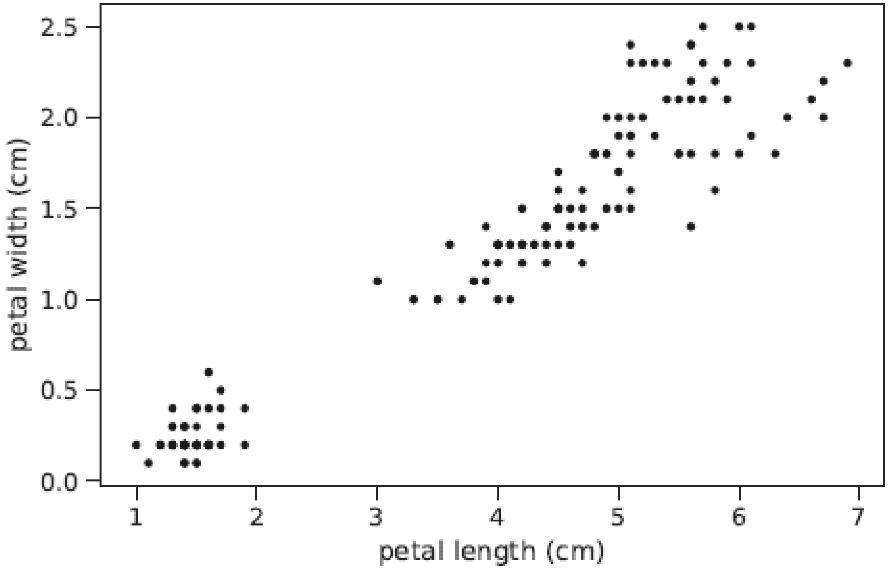
Unsupervised Learning

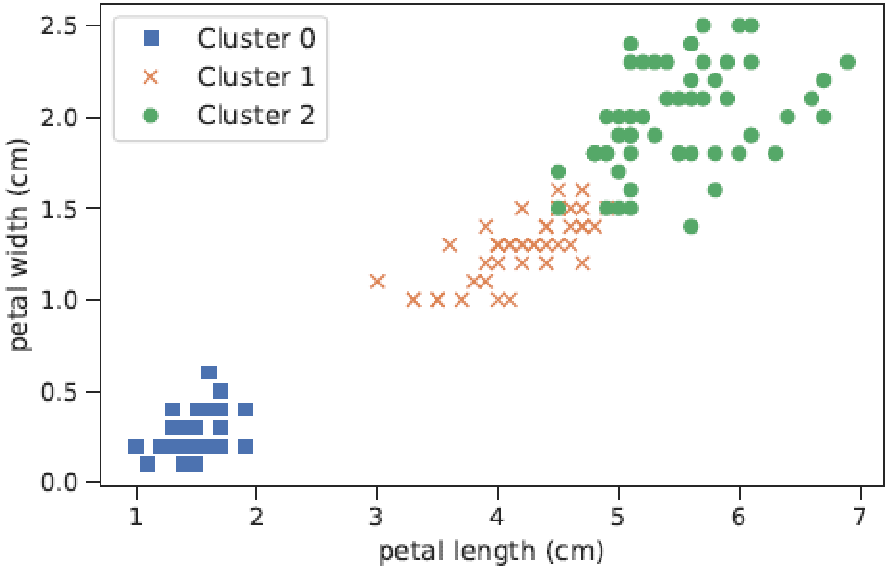
Supervised Learning	**Unsupervised Learning**	Reinforcement Learning
Labelled data with guidance	No labels	Interacts with environment, decides action, learns by trial and error method

Unsupervised Learning - Clustering

Goal:

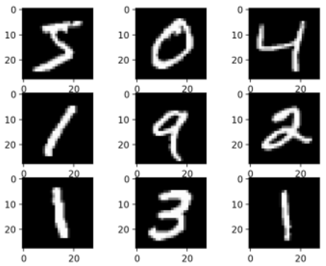
Partition the input into regions that contain “similar” points.

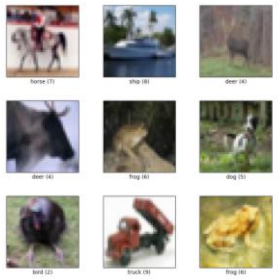




Unsupervised Learning - Clustering

$$\text{high-dimensional output } \boldsymbol{x}_n \in \mathbb{R}^D$$





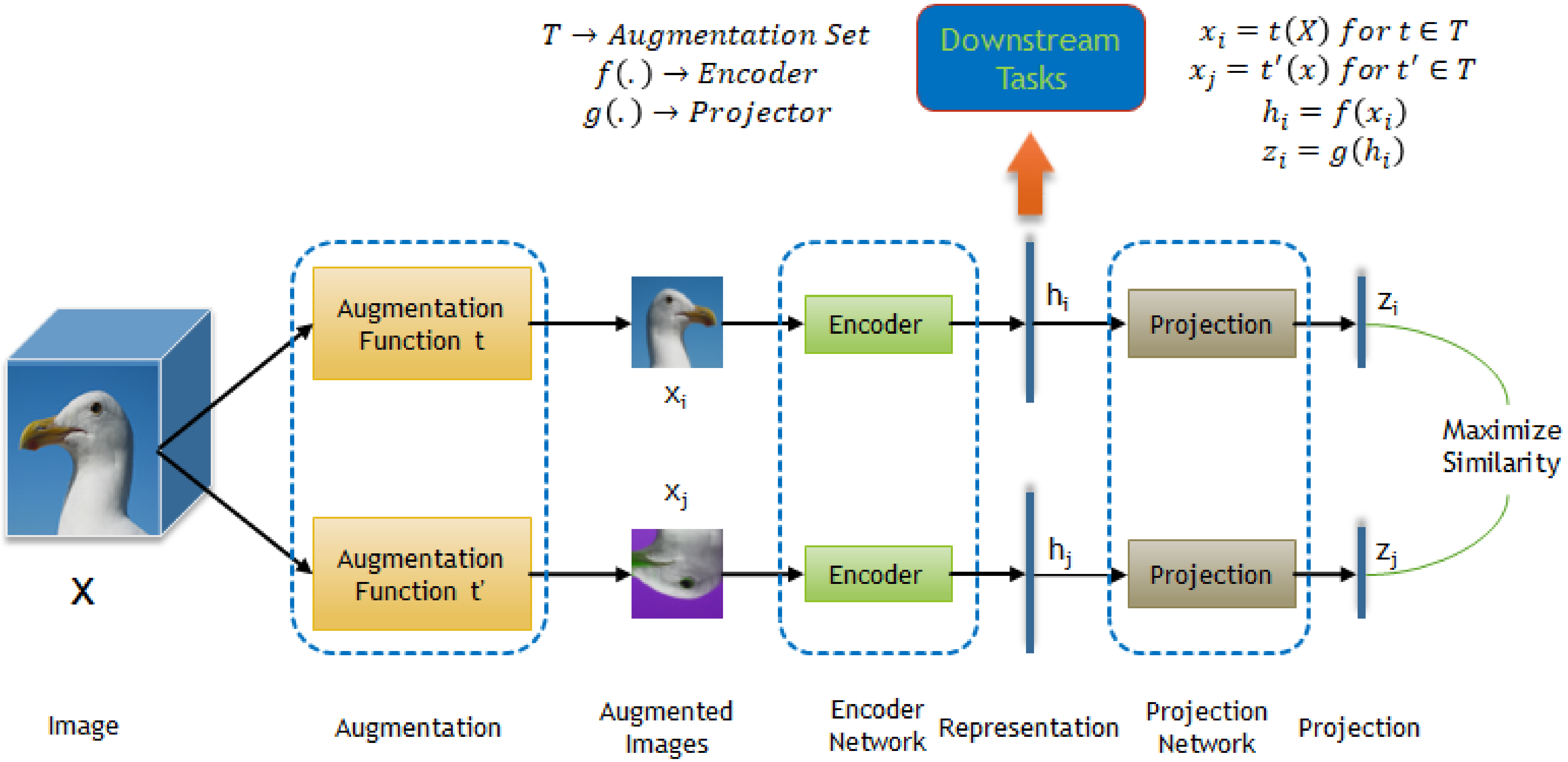
$$\textbf{latent factors } \boldsymbol{z}_n \in \mathbb{R}^K$$

$$\boldsymbol{z}_n \rightarrow \boldsymbol{x}_n$$

Linear Model $p(\boldsymbol{x}_n | \boldsymbol{z}_n; \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_n | \mathbf{W}\boldsymbol{z}_n + \boldsymbol{\mu}, \boldsymbol{\Sigma})$

Non-linear Model $p(\boldsymbol{x}_n | \boldsymbol{z}_n; \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_n | f(\boldsymbol{z}_n; \boldsymbol{\theta}), \sigma^2 \mathbf{I})$

Unsupervised Learning – Self-supervised Learning



From https://medium.com/analytics-vidhya/self-supervised-representation-learning-in-computer-vision-part-2-8254aaee937c

Evaluation

TRAIN → **VALIDATE** → **TEST**

CROSS VALIDATE

DO NOT MIXUP TRAINING, VALIDATION AND TEST DATA

Reinforcement Learning

Supervised Learning	Unsupervised Learning	**Reinforcement Learning**
Labelled data with guidance	No labels	Interacts with environment, decides action, learns by trial and error method

Reinforcement Learning

A system or agent has to learn how to interact with its environment.

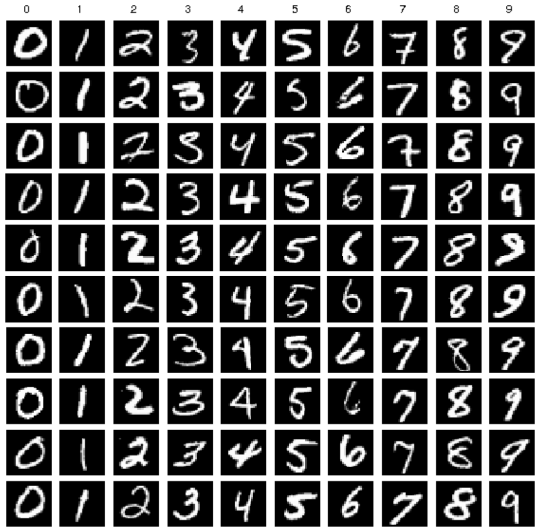
This can be encoded by means of a **policy a = (x)**, which specifies which action to take in response to each possible **input x** (derived from the environment state).

(a) (b)

Figure 1.10: Examples of some control problems. (a) Space Invaders Atari game. From https://gym.openai.com/envs/SpaceInvaders-v0/. (b) Controlling a humanoid robot in the MuJuCo simulator so it walks as fast as possible without falling over. From https://gym.openai.com/envs/Humanoid-v2/.

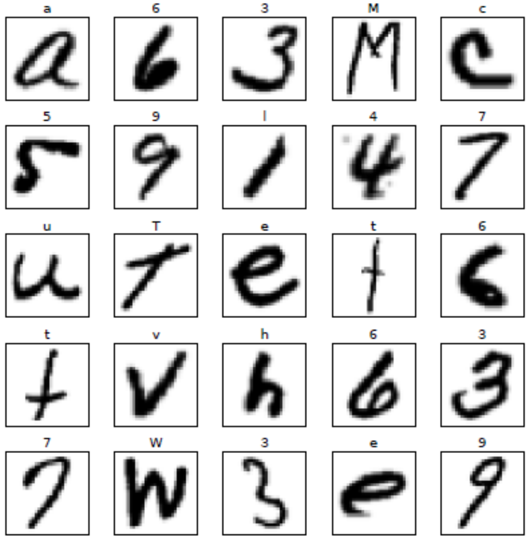
Common Small Image Datasets

MNIST

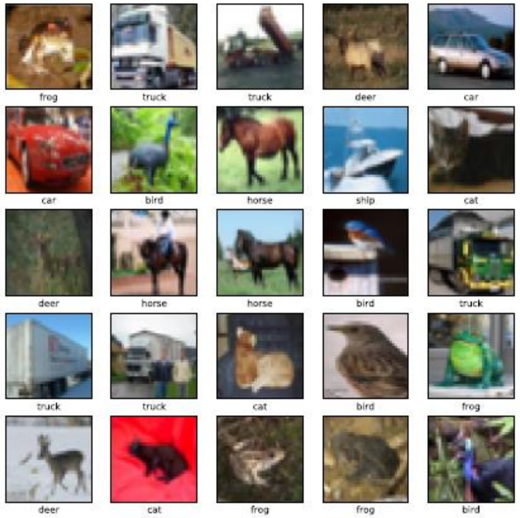


Fashion - MNIST

EMNIST



CIFAR



Common Large Image Datasets

ImageNet

- This dataset spans 1000 object classes
- 1,281,167 training images,
- 50,000 validation images and
- 100,000 test images

https://www.image-net.org/download.php

Discuss pros and cons of this dataset.

Natural Language Processing

IMDB movie review

```
data.shape
```

```
(100000, 5)
```

```
data.head()
```

	Unnamed: 0	type	review	label	file
0	0	test	Once again Mr. Costner has dragged out a movie...	neg	0_2.txt
1	1	test	This is an example of why the majority of acti...	neg	10000_4.txt
2	2	test	First of all I hate those moronic rappers, who...	neg	10001_1.txt
3	3	test	Not even the Beatles could write songs everyon...	neg	10002_3.txt
4	4	test	Brass pictures (movies is not a fitting word f...	neg	10003_3.txt

Natural Language Processing (NLP)

- **Classification**

Natural Language Processing

Natural Language Processing (NLP)

- **Translation**
  - ✓ Canadian parliament (English-French pairs)
  - ✓ the European Union (Europarl).

- **Document summarization, Question answering**

T: In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

Q1: What causes precipitation to fall? A1: **gravity**
Q2: What is another main form of precipitation besides drizzle, rain, snow, sleet and hail? A2: **graupel**
Q3: Where do water droplets collide with ice crystals to form precipitation? A3: **within a cloud**

Table 1.4: Question-answer pairs for a sample passage in the SQuAD dataset. Each of the answers is a segment of text from the passage. This can be solved using sentence pair tagging. The input is the paragraph text T and the question Q. The output is a tagging of the relevant words in T that answer the question in Q. From Figure 1 of [Raj+16]. Used with kind permission of Percy Liang.