

Data ■ Model ■ Loss

An open resource from the Aalto Dictionary of ML

Alexander Jung

October 7, 2025

Licensed under CC BY 4.0



At a glance — goals

- ▶ Understand the three components of **machine learning** (ML): **data**, **model**, **loss**.
- ▶ Identify **features** vs **labels** in common modalities (image, audio).
- ▶ Explain how a **hypothesis** maps $\mathcal{X} \rightarrow \mathcal{Y}$.
- ▶ Compare common **loss functions** and when to use them.

Data point = An Image z



Features:

- ▶ x_1, \dots, x_d : Colour intensities of all image pixels.
- ▶ x_{d+1} : Time-stamp of the image capture.
- ▶ x_{d+2} : Spatial location of the image capture.

Labels:

- ▶ y_1 : Number of cows depicted.
- ▶ y_2 : Number of wolves depicted.
- ▶ y_3 : Condition of the pasture (e.g., healthy, overgrazed).

Data point = An Audio Recording \mathbf{z}

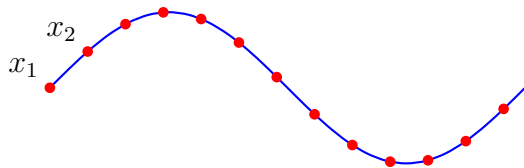
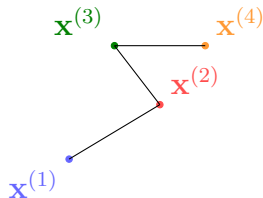
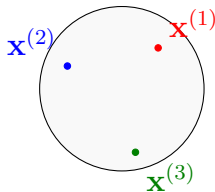
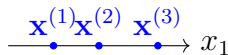


Figure: An audio signal (blue waveform) \mathbf{z} and its discretized signal samples (red dots) which can be used as its **features** x_1, \dots, x_d .

Feature space

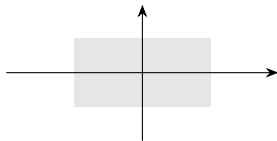
- ▶ often we use a fixed number $d \in \mathbb{N}$ of **features**
- ▶ stack them into a **feature vector** $\mathbf{x} = (x_1, \dots, x_d)$
- ▶ **feature vectors** belong to some **feature space** \mathcal{X}
- ▶ most widely-used (by far) choice is $\mathcal{X} = \mathbb{R}^d$



Label space



(a) $\mathcal{Y} = \mathbb{R}$ (regression)

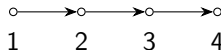


(b) $\mathcal{Y} = \mathbb{R}^2$ (multi-label regression)

•
"hot"

•
"cold"

(c) $\mathcal{Y} = \{y_1, y_2\}$ (binary classification)



(d) $\mathcal{Y} = \{1, 2, 3, 4\}$ (ordinal regression)

Figure: Examples of label spaces and corresponding ML flavours.

Goal of ML: Predict Label from Features

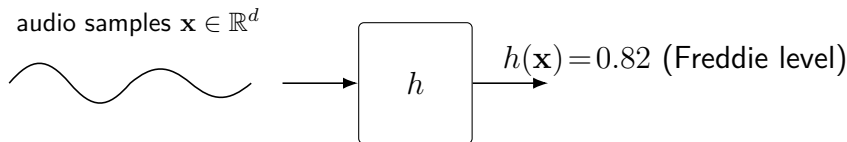
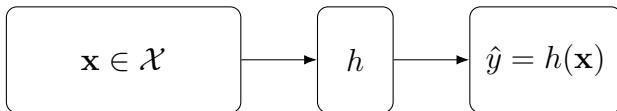


Figure: A hypothesis $h : \mathcal{X} \rightarrow \mathcal{Y}$ maps the **features** $\mathbf{x} \in \mathcal{X}$ of a **data point** to a **prediction** $h(\mathbf{x}) \in \mathcal{Y}$ of the **label**. For example, the **ML** application <https://freddiemeter.withyoutube.com/> uses the samples of an audio recording as **features** predict how closely a person's singing resembles that of Freddie Mercury.

From Features to Prediction



Core Problem of ML given **data** $\{(\mathbf{x}^{(r)}, y^{(r)})\}_{r=1}^m$ and **model** \mathcal{H} , learn (or find) $\hat{h} \in \mathcal{H}$ such that $\hat{h}(\mathbf{x}) \approx y$ for any **data point** with **features** \mathbf{x} and **label** y .

Model = A Set of Hypothesis Maps

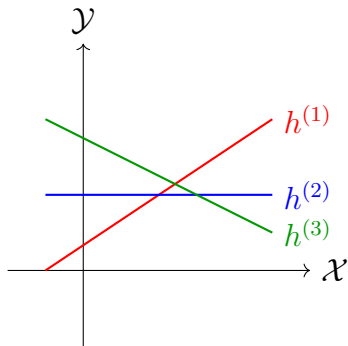
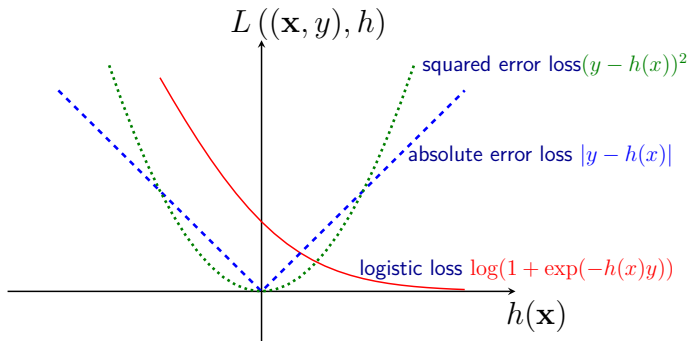


Figure: A hypothesis space $\mathcal{H} = \{h^{(1)}, h^{(2)}, h^{(3)}\}$ consisting of three linear maps.

Which one of the hypothesis maps is the best?

Loss function



A **loss function** $L((\mathbf{x}, y), h)$ measures the error (or “**loss**”), incurred by predicting the **label** y of a **data point** with **feature vector** \mathbf{x} .

Which Loss function should we use?

The choice of **loss function** influences

- ▶ computational aspects,
- ▶ statistical aspects (robustness, generalization, ...), and
- ▶ interpretability

of the resulting **ML** method.

Reuse & Citation

This deck is part of the **Aalto Dictionary of Machine Learning** (open source).

License: **CC BY 4.0**. Please cite: *Jung, A. et.al. (2025). The Aalto Dictionary of Machine Learning*. DOI: 10.5281/zenodo.17273736.

Source and updates: <https://github.com/AaltoDictionaryofML/AaltoDictionaryofML.github.io>