Machine Learning

<u>def</u> Program learns from experience, E, with respect to class of tasks T, and performance measure, P, if its performance at T improves by P w/ E.

<u>rem</u>
- E : data (training)
- P : loss (test), reward
- T : classification, regression, expert selection.
  parameters

<u>defs</u> Given a loss $l_i(\theta)$ for datum/instance i <u>risk</u> is
$R(\theta) = \mathbb{E}\, l_i(\theta)$ $\left(= \mathbb{E}\, l_1(\theta) \text{ if id.}\right)$ and <u>empirical risk</u> is

$$R_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} l_i(\theta)$$

<u>ex</u> Supervised learning

$$\begin{tikzcd} & X \arrow[d] \\ \theta \arrow[r] & Y \end{tikzcd}$$

Want to learn $\mathbb{E}\, Y|X$ from iid samples $\{x_i, y_i\}$ (parametrized by θ)

Typically: $x_i \in \mathbb{R}^p$ $y_i \in \mathbb{R}$ (or $\{-1,+1\}$)

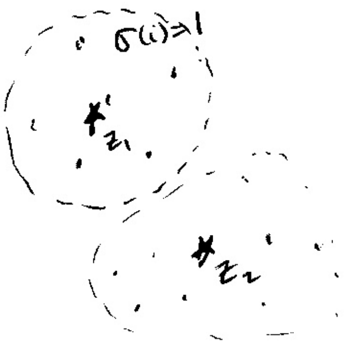
0-1 loss: $l_{0\text{-}1}(g_\theta(X), Y) = \mathbb{1}\{g_\theta(X) \neq Y\}$ $\quad g_\theta : \mathbb{R}^p \to \mathbb{R}$

0-1 risk: $\mathbb{E}\, l_{0\text{-}1}(g_\theta(X), Y) = \mathbb{P}\{g_\theta(X) \neq Y\}$

<u>ex</u> Unsupervised learning

$$\Theta \longrightarrow X$$

Want to summarize/compress/learn distribution of X. Clustering for example is



$\Theta = \left\{\{z_k\}_{k=1}^{c} \subseteq \mathbb{R}^p,\ \sigma : \{1,\dots,n\} \to \{1,\dots,c\}\right\}$ and loss may be (distortion)

$$l_i(\Theta) = \|x_i - z_{\sigma(i)}\|_2^2 = \sum_j (x_{ij} - z_{\sigma(i)j})^2$$