Statistical learning theory and the "Probably Approximately Correct" (PAC) framework*

Notation

given to us	$\mathbf{z} = (\mathbf{x}, y)$	$\mathbf{x} \in \mathcal{X}$	instance/features (e.g., "picture" or actual pixel values)	
	$\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$	$y \in \mathcal{Y}$	labels, e.g., *binary classification* uses $\mathcal{Y} = \{0,1\}$ or $\{-1,1\}$	
		$\mathbf{z}_i \sim \mathcal{D}$	observed realizations	
		$S = \left(\mathbf{z}_1, \ldots, \mathbf{z}_m\right)$	data set, $m = \#$ observations	
we choose		$h : \mathcal{X} \to \mathcal{Y}$	classifier, often parameterized by $\mathbf{w} \in \mathbb{R}^d$	
		$\mathcal{H}$	set of classifiers, $h \in \mathcal{H}$	
metrics		$\ell : \mathcal{H} \times \mathcal{Z} \to \mathbb{R}^+$	loss function, usually non-negative, often bounded or Lipschitz	
		$L_{\mathcal{D}}(h) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{z} \sim \mathcal{D}}\ \ell(h, \mathbf{z})$	true **risk** (this is what we care about)	same as "testing" error [theoretical]
		$\widehat{L}_S(h) \stackrel{\text{def}}{=} \dfrac{1}{m} \sum_{i=1}^{m} \ell(h, \mathbf{z}_i)$	empirical **risk**	same as "training" error [observable]

* This is the standard setup for **supervised, passive, statistical, batch** learning.
There are many other interesting setups, such as **unsupervised, active, adversarial, online** learning,
but our setup is sufficient to convey the main issues