PROBABILITY DISTRIBUTION D on $\mathcal{X} \times \mathcal{Y}$

data

$(\vec{x}_1, y_1), (\vec{x}_2, y_2), \ldots$

$\vec{x} \in \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix} \in \mathbb{R}^{d+1}$

training data

$(\vec{x}_1, y_1), (\vec{x}_2, y_2), \ldots$

test data

$(\vec{x}_{t_1}, y_{t_1}), \ldots (\vec{x}_{t_n}, y_{t_n})$

HYPOTHESIS/MODEL SET

$\mathcal{H}$

$h_{\vec{w}}(\vec{x}) = \text{sign}(\langle \vec{w}, \vec{x} \rangle)$

ALGORITHM

$\mathcal{A}$

perceptron (SGD)

LOSS FUNCTION

$\ell()$

$\ell(h_{\vec{w}}, (\vec{x}, y)) =$

$= \max\{0, -y\langle \vec{w}, \vec{x} \rangle\}$

ERM: pick $\vec{w}^*$ that gives the minimum training error

FINAL HYPOTHESS

$h_{\vec{w}^*}$ (fix $\vec{w}^*$)

estimate the generalization error for $h_{\vec{w}^*}$ ($L_D(h_{\vec{w}^*})$