ERM $\longleftrightarrow$ MAX LIKELIHOOD

MAX-LIKELIHOOD?

$D = \{(x_i, y_i)\}_{i=1,\dots,m}$ $\quad D \sim p^m$ $\quad , p = p(x, y)$

$$p(x,y) = p(x)\underbrace{p(y|x)}_{\text{we make hypothesis on it}}$$

$p(y|x) = p(y|x, \vartheta)$

$$\mathcal{L}(\vartheta; D) = \sum_{i=1}^{m} \log p(y_i | x_i, \vartheta)$$

$$\vartheta_{ML} = \arg\max_{\vartheta} \mathcal{L}(\vartheta; D) = \arg\min_{\vartheta} -\mathcal{L}(\vartheta; D) =$$

cross entropy

$$\arg\min -\mathcal{L}(\vartheta; D) = \arg\min_{\vartheta} \left(-\frac{1}{m} \sum_{i=1}^{m} \log p(y_i | x_i, \vartheta) \right)$$

$$\simeq \arg\min_{\vartheta} \mathbb{E}_{p(x,y)} \left[-\log p(y | x, \vartheta) \right]$$

$$\Rightarrow \quad \underline{\ell(x, y, \vartheta) = -\log p(y \mid x, \vartheta)}$$

What is $\mathcal{H}$?

- REGRESSION $\quad h(x, \vartheta) = \mathbb{E}_y\left[p(y \mid x, \vartheta)\right]$
- CLASSIFICATION $\quad h(x, \vartheta) = \arg\max_y p(y \mid x, \vartheta)$ (Bayes decision rule)

Choices of $p(y \mid x, \vartheta)$

- REGRESSION $\quad p(y \mid x, \vartheta) = \mathcal{N}(h_\vartheta(x), \sigma^2)$

$$\Rightarrow -\log p(y \mid x, \vartheta) \propto (y - h_\vartheta(x))^2$$