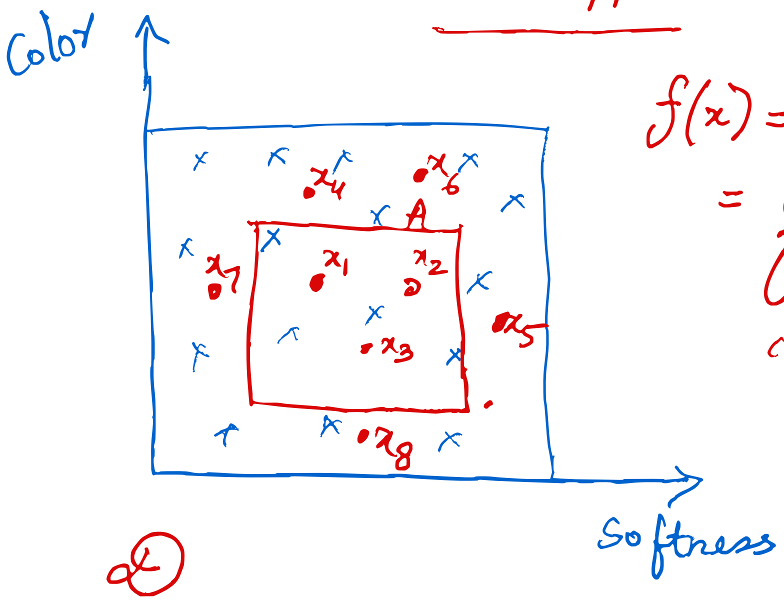
Probably

With probability $> 1-\delta$ over samples $S$, the generalization risk $R(h_S)$ of an ERM rule $h_S$

$$\underbrace{R(h_S) < \underbrace{\frac{1}{m}\log\left(\frac{|\mathcal{H}|}{\delta}\right)}_{\text{approximately}}}$$

Proof: Next time




$$f(x) = y = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \text{ (not tasty)} \end{cases}$$

$$\boxed{\text{ERM} \quad h(x) = \begin{cases} 1 & \text{if } x = x_1, x_2, x_3 \\ 0 & \text{o.w.} \end{cases}}$$

Overfitting

$$\underset{h \in \mathcal{H}}{\arg\min}\ \hat{R}_S(h) \qquad (\text{ERM})$$

Any function $h \in \mathcal{H}$ has the form

$$h(x) = \begin{cases} 1 & \text{if } x \in \text{rectangle similar to } A \\ 0 & \text{o.w.} \end{cases}$$

Inductive bias to avoid overfitting

PAC learning