

$$\Delta P[|E_{\text{tr}}(h) - E(h)| > \epsilon] \leq 2e^{-2\epsilon^2 N}$$

$\underbrace{\quad}_{\text{training error}}$
 $\underbrace{\quad}_{\text{transition error?}}$

Δ also, when we have multiple hypothesis:
 we want an upper bound

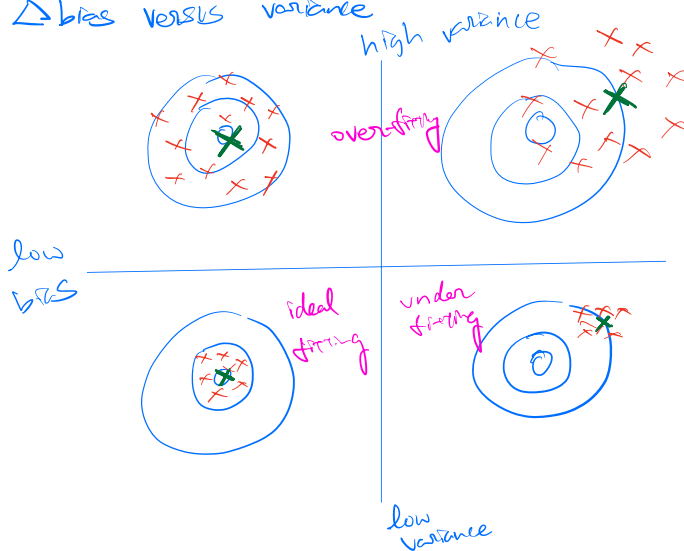
$$P[\sup_{h \in \mathcal{H}} |E_{\text{tr}}(h) - E(h)| > \epsilon]$$

Δ How many training example do you need.

$$N \geq \frac{1}{2\epsilon^2} \log \frac{2/K}{\delta} \quad \text{where } \delta = 2/K e^{-2\epsilon^2 N}$$

$$E_{\text{tr}}(h^*) - r \leq E(h^*)$$

Δ bias versus variance



$$E_{\text{tr}}(h^*) - r$$

large E_{train}
 large E_{test}

overfitting. \Rightarrow high variance
 sensitive to noise

$$\Delta H = w^T x + b$$

$|K| \rightarrow$ is actually infinity
 (non-terminal)

How do we shrink it?
 (improve)

Δ on P17 - LF.

h_1 & h_2 although
 mathematical
 different

it is empirically same.

