Model Selection and Validation

Model Selection: Choosing best Alg. & Parameters.

Example

Fitting a polynomial to learn function $\mathbb{R} \to \mathbb{R}$

§2 Validation

Theorem 1

Let h be a predictor and $V = (x_1, y_1), \ldots, (x_{m_v}, y_{m_v})$ a validation set sampled $\sim \mathcal{D}$.
Assume our loss function $\in [0,1]$.
Now for every $\delta \in (0,1)$:
With prob. of at least $1-\delta$, we have

$$|L_V(h) - L_\mathcal{D}(h)| \leq \sqrt{\frac{\log(2/\delta)}{2m_v}}$$

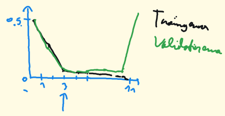
Theorem 2

Let $H = \{h_1, \ldots, h_r\}$. Other Assumptions as above
with prob. at least $1-\delta$

$$\forall h \in H: \ |L_\mathcal{D}(h) - L_V(h)| \leq \sqrt{\log\left(\frac{2|H|/\delta}{2m_v}\right)}$$

§ Model Selection curve

Plot training & Validation error against complexity of the model
e.g. polynomials



§3 What to do if learning fails

- Get a larger set of samples
- Change hypothesis class
  - enlarge it
  - reduce it
  - completely change it
  - change parameters considered
- Change the feature representation
- Change the optimization algorithm

§3.1 Error decomposition

h_S is learned on training set S
$h^* = \operatorname{argmin}_{h \in H} L_\mathcal{D}(h)$

$$L_\mathcal{D}(h_S) = \underbrace{L_\mathcal{D}(h^*)}_{\text{approx. error}} + \underbrace{\left(L_\mathcal{D}(h_S) - L_S(h^*)\right)}_{\text{estimation error}}$$

$$L_\mathcal{D}(h_S) = \underbrace{\left(L_\mathcal{D}(h_S) - L_V(h_S)\right)}_{\text{tightly bound}} + \underbrace{\left(L_V(h_S) - L_S(h_S)\right)}_{\text{measure for overfitting}} + \underbrace{L_S(h_S)}_{\text{measure for underfitting}}$$

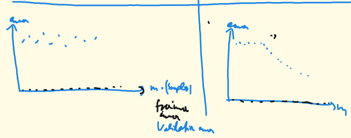
$$L_S(h_S) = \underbrace{\left(L_S(h_S) - L_S(h^*)\right)}_{\leq 0} + \underbrace{\left(L_S(h^*) - L_\mathcal{D}(h^*)\right)}_{\substack{h^* \text{ independent of } S \\ \text{tightly bounded}}} + \underbrace{L_\mathcal{D}(h^*)}_{\text{the approx. error}}$$

if $L_S(h_S)$ is large $\Rightarrow$ approx. error is large
$\not\Leftarrow$

§3.2 Learning Curves

Assume $L_S(h_S)$ is small $(=0)$

$m < VC$ dim of the class high approximation error	$m > 2 \cdot VC$ dim approximation error is 0



3.3 Summary