More about Validation Set

I used a validation set, but my model still overfitted?
Validation Set

Training Set $\mathcal{D}_{train}$

Model $\mathcal{H}_1$  \[ h_1^* = \arg\min_{h \in \mathcal{H}_1} L(h, \mathcal{D}_{train}) \]

Model $\mathcal{H}_2$  \[ h_2^* = \arg\min_{h \in \mathcal{H}_2} L(h, \mathcal{D}_{train}) \]

Model $\mathcal{H}_3$  \[ h_3^* = \arg\min_{h \in \mathcal{H}_3} L(h, \mathcal{D}_{train}) \]

Validation Set $\mathcal{D}_{val}$

- $L(h_1^*, \mathcal{D}_{val}) = 0.9$
- $L(h_2^*, \mathcal{D}_{val}) = 0.7$
- $L(h_3^*, \mathcal{D}_{val}) = 0.5$

Testing Set $\mathcal{D}_{test}$

Approximation of $\mathcal{D}_{all}$
Training Set $\mathcal{D}_{\text{train}}$

Model $\mathcal{H}_1$  \[ h_1^* = \arg \min_{h \in \mathcal{H}_1} L(h, \mathcal{D}_{\text{train}}) \]

Model $\mathcal{H}_2$  \[ h_2^* = \arg \min_{h \in \mathcal{H}_2} L(h, \mathcal{D}_{\text{train}}) \]

Model $\mathcal{H}_3$  \[ h_3^* = \arg \min_{h \in \mathcal{H}_3} L(h, \mathcal{D}_{\text{train}}) \]

Validation Set $\mathcal{D}_{\text{val}}$

\[ L(h_1^*, \mathcal{D}_{\text{val}}) = 0.9 \]

\[ L(h_2^*, \mathcal{D}_{\text{val}}) = 0.7 \]

\[ L(h_3^*, \mathcal{D}_{\text{val}}) = 0.5 \]

$\mathcal{H}_{\text{val}} = \{h_1^*, h_2^*, h_3^*\}$ \[ h^* = \arg \min_{h \in \mathcal{H}_{\text{val}}} L(h, \mathcal{D}_{\text{val}}) \]

Using validation set to select model = considered as “training” by $\mathcal{D}_{\text{val}}$

Your model is \[ \mathcal{H}_{\text{val}} = \{h_1^*, h_2^*, h_3^*\} \]
Using validation set to select model = considered as “training” by $\mathcal{D}_{val}$

Your model is $\mathcal{H}_{val} = \{h_1^*, h_2^*, h_3^*\}$

$L(h^{train}, \mathcal{D}_{all}) - L(h^{all}, \mathcal{D}_{all}) \leq \delta$

$P(\mathcal{D}_{train} \text{ is bad}) \leq |\mathcal{H}| \cdot 2\exp(-2N\varepsilon^2)$

$L(h^{val}, \mathcal{D}_{all}) - L(h^{all}, \mathcal{D}_{all}) \leq \delta$

$P(\mathcal{D}_{val} \text{ is bad}) \leq |\mathcal{H}_{val}| \cdot 2\exp(-2N_{val}\varepsilon^2)$

It is small.

Hopefully ...... 😊