

Limits of Learning

<https://piazza.com/class/k5pgep4orjm7ae?cid=8>

CIML2 Bayes-Optimal Classifier

(What about regression?)

Conditional Expectation:

$$h(\mathbf{x}) = \mathbb{E}[t|\mathbf{x}] = \int tp(t|\mathbf{x}) dt.$$

CIML2 Inductive Biases

CIML2 Parameters vs Hyperparameters

CIML2 Training Data/Testing Data

CIML2 Overfitting/Underfitting

PRML 3.2 Bias-Variance Decomposition

Only for Squared Loss: $h(\mathbf{x}) = \mathbb{E}[t|\mathbf{x}]$

PRML 1.5.5:
$$\mathbb{E}[L] = \int \{y(\mathbf{x}) - \mathbb{E}[t|\mathbf{x}]\}^2 p(\mathbf{x}) d\mathbf{x} + \int \{\mathbb{E}[t|\mathbf{x}] - t\}^2 p(\mathbf{x}) d\mathbf{x}. \quad (1.90)$$

Over all possible datasets \mathcal{D} $\{y(\mathbf{x}; \mathcal{D}) - h(\mathbf{x})\}^2$.

$$\begin{aligned} & \{y(\mathbf{x}; \mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x}; \mathcal{D})] + \mathbb{E}_{\mathcal{D}}[y(\mathbf{x}; \mathcal{D})] - h(\mathbf{x})\}^2 \\ &= \{y(\mathbf{x}; \mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x}; \mathcal{D})]\}^2 + \{\mathbb{E}_{\mathcal{D}}[y(\mathbf{x}; \mathcal{D})] - h(\mathbf{x})\}^2 \\ & \quad + 2\{y(\mathbf{x}; \mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x}; \mathcal{D})]\}\{\mathbb{E}_{\mathcal{D}}[y(\mathbf{x}; \mathcal{D})] - h(\mathbf{x})\}. \\ & \mathbb{E}_{\mathcal{D}} [\{y(\mathbf{x}; \mathcal{D}) - h(\mathbf{x})\}^2] \\ &= \underbrace{\{\mathbb{E}_{\mathcal{D}}[y(\mathbf{x}; \mathcal{D})] - h(\mathbf{x})\}^2}_{(\text{bias})^2} + \underbrace{\mathbb{E}_{\mathcal{D}} [\{y(\mathbf{x}; \mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x}; \mathcal{D})]\}^2]}_{\text{variance}}. \quad (3.40) \end{aligned}$$

expected loss = (bias)² + variance + noise

**Next module: nearest neighbors and curse of
dimensionality**

PRML 1.4, CIML 3