Supervised Learning: Evaluating Errors

Question 1:
How can we estimate the model performance properly for unknown data?

Question 2:
How can we choose the optimal hyperparameters?
Supervised Learning: Evaluating Errors

Solution

Divide data \( D = \{(x_1, t_1), \ldots, (x_N, t_N)\} \) in 3 groups:

- **Training set** \( D_{\text{train}} (\pm \Theta \% \text{ of } D) \):
  - Minimize the error \( E(y(x), t) \) for \((x, t) \in D_{\text{train}}\)

- **Validation set** \( D_{\text{val}} (\pm 10 \% \text{ of } D) \):
  - Used to estimate test error \( E(y(x_{\text{val}}, w^*), t_{\text{val}}) \)

- **Test set** \( D_{\text{test}} (\pm 10 \% \text{ of } D) \):
  - **final test/generalization error estimate** \( E(y(x_{\text{test}}, w^*), t_{\text{test}}) \)

Can never be part of model selection!!!
Supervised Learning: Small Datasets

- Small dataset ➔ small validation and test set
- Approximate validation step!

**Cross-validation**

- Split data: \( D = \{(x_1, t_1), \ldots, (x_N, t_N)\} \) into \( K \)-folds
- Train \( y \) on \( K-1 \) folds \( \hat{y}^{\text{K-1}}(x) \)

![K-fold splitting of dataset](ESL 7.10)

\[
\begin{array}{ccccc}
1 & 2 & 3 & 4 & 5 \\
Train & Train & Validation & Train & Train \\
\end{array}
\]

**Figure:** K-fold splitting of dataset (ESL 7.10)

- \( K = N \) : leave-one-out cross validation
Cross-Validation

- \( K \) trained functions \( \hat{y}^{-k}(x) \)

- Indexing function \( \kappa : \{1, \ldots, N\} \mapsto \{1, \ldots, k\} \)

- Estimate of prediction error

\[
CV(\hat{y}) = \frac{1}{N} \sum_{i=1}^{N} E(\hat{y}^{-\kappa(i)}(x_i), t)
\]

2 tasks

1. model selection (optimal hyperparameters) \( \hat{y}^{-k} \)
2. estimate model performance \( \hat{y}^{-k} \)
Cross-Validation: Model Selection

- Hyperparameter selection $\alpha$
  
  $CV(\hat{y}_\alpha) = \frac{1}{N} \sum_{h=1}^{N} E(\hat{y}_{\alpha}^h(x_h), t_h)$

- Optimal $\alpha^* = \arg\min_{\alpha} CV(\hat{y}_\alpha)$

- Multiple hyperparameters: $\alpha \in \{\alpha_1, \alpha_2\}, \beta \in \{\beta_1, \beta_2, \beta_3\}$

  - How many times should CV be performed? $2 \times 3$
  - Total number of training runs? $6K$
Cross-Validation: Test Error Estimation

- After Model selection $\alpha^*, \beta^*$
- Retrain $f$ on all K folds with $\alpha^*, \beta^*$
- Evaluate model on held-out test set
- Nested cross validation!
Nested Cross-Validation

Figure: Nested cross-validation