svm_ward2025

This program is parametrised by a classification dataset, with d features, split according to K-fold cross-validation. The superscript k is used to denote fold k and bars over the symbols are used to distinguish validation from training.

	`Training`	`Validation`	
`Number of examples:`	n^k	$\bar{n}^k$	$\in \mathbb{N}$
`Feature vectors:`	x_i^k	$\bar{x}_i^k$	$\in \mathbb{R}^d$
`Labels:`	y_i^k	$\bar{y}_i^k$	$\in \{-1, +1\}$

The Radial Basis Function (RBF) kernel matrix gives a distance between every two examples j and j in fold k defined by

$$Q(\gamma)_{ij}^k := y_i^k y_j^k \exp\left(-\gamma \|x_i^k - x_j^k\|^2\right). \tag{RBF}$$

The upper-level program is to choose hyperparameters C and γ that minimise the average over K folds of validation hinge loss error.

$$\begin{aligned}
\underset{C,\gamma,\zeta,\alpha}{\text{minimise}} \quad & \sum_{k=1}^{K} \frac{1}{\bar{n}^k} \sum_{i=1}^{\bar{n}^k} \zeta_i^k \\
\text{subject to} \quad & C \geq 0, \\
& \gamma \geq 0, \\
& \left.\begin{aligned} &\zeta_i^k \geq 0 \\ &\zeta_i^k \geq 1 - \sum_{j=1}^{n^k} \alpha_j^k \bar{Q}(\gamma)_{ij}^k - \bar{y}_i^k b^k \end{aligned}\right\} \begin{aligned} &\text{for } k = 1, \ldots, K, \\ &\quad \text{for } i = 1, \ldots, \bar{n}^k, \end{aligned} \\
& \alpha^k \text{ solve (Dual-SVM)} \quad \text{for } k = 1, \ldots, K \quad \text{for } i = 1, \ldots, n^k.
\end{aligned}$$

The lower level problem is the Dual Support Vector Machine which aims to minimise the hinge loss training error while maximising the width of the margin.

$$\begin{aligned}
\underset{\alpha^k \in \mathbb{R}^{n^k}}{\text{minimise}} \quad & \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i^k \alpha_j^k Q(\gamma)_{ij}^k - \sum_{i=1}^{n} \alpha_i \\
\text{subject to} \quad & 0 \leq \alpha_i^k \leq C \quad \text{for } i = 1, \ldots, n, \\
& \sum_{i=1}^{n} \alpha_i^k y_i = 0,
\end{aligned} \tag{Dual-SVM}$$

For further details see [1, Section 4].

References

[1] Samuel Ward, Alain Zemkoho, and Selin Ahipasaoglu. Mathematical programs with complementarity constraints and application to hyperparameter tuning for nonlinear support vector machines, 2025.