svm_linear

Given a dataset, we split it into n training examples and $\bar{n}$ validation examples. Denote by $x_i \in \mathbb{R}^d$ the feature vector and $y_i \in \{-1, +1\}$ the label of the ith *training* example. Using a bar over the symbols, denote $\bar{x}_i \in \mathbb{R}^d$ the feature vector and $\bar{y}_i \in \{-1, +1\}$ the ith *validation* example.

The upper-level problem is to choose the regularisation hyperparameter C and slack variables ζ to minimise the validation loss.

$$\begin{aligned} \underset{C,\zeta,w,\xi,b}{\text{minimise}} \quad & \sum_{i=1}^{\bar{n}} \zeta_i^k \\ \text{subject to} \quad & C \geq 0, \\ & \zeta_i \geq 0 && \text{for } i = 1, \dots, \bar{n}, \\ & \zeta_i \geq 1 - \bar{y}_i \left(w^\top \bar{x}_i + b \right) && \text{for } i = 1, \dots, \bar{n}, \\ & w, b \text{ solves (SVM)} \end{aligned}$$

The lower-level problem is to choose weights w, bias b and slack variables ξ to minimise the training loss plus regularisation.

$$\begin{aligned} \underset{w \in \mathbb{R}^d, \xi \in \mathbb{R}^n, b \in \mathbb{R}}{\text{minimise}} \quad & \frac{1}{2} w^\top w + C \frac{1}{n} \sum_{i=1}^{n} \xi_i \\ \text{subject to} \quad & \xi_i \geq 1 - y_i \left(w^\top x_i + b \right) && \text{for } i = 1, \dots, n, \\ & \xi_i \geq 0 && \text{for } i = 1, \dots, n. \end{aligned} \tag{SVM}$$