Supplementary Material: An LP for Sequential Learning Under Budget Constraints

Joseph Wang
Boston University

Kirill Trapeznikov
Boston University

Venkatesh Saligrama
Boston University

]

Additional Details for Proof of Theorem 3.1 The empirical risk, as shown in the Proof of Theorem 3.1, can be expressed:

$$R(f^1,\ldots,f^K,x,y) = 1+\alpha\sum_{k=1}^{K} c_k - \sum_{k=1}^{K}\pi^k + \sum_{k=1}^{K}\pi^k \max\left(\mathbb{1}_{g^k(x^k)>0}, \mathbb{1}_{g^1(x^1)\le 0},\ldots,\mathbb{1}_{g^{k-1}(x^{k-1})\le 0}\right).$$

We focus solely on the maximization term, which can be equivalently expressed by introducing new variables $\lambda_1^1,\ldots,\lambda_K^K$:

$$\sum_{k=1}^{K}\pi^k \max_{\lambda_k^1,\ldots,\lambda_k^k\in[0,1]}\left(\lambda_k^k \mathbb{1}_{g^k(x^k)>0} + \sum_{j=1}^{k-1}\lambda_k^j \mathbb{1}_{g^j(x^j)\le 0}\right)$$

where the variables $\lambda_1^1,\ldots,\lambda_K^K$ are constrained $\sum_{j=1}^{k}\lambda_k^j = 1$. Consider the first j such that $g^j(x^j)\le 0$. For all $k>j$, one optimal solution for the indicators is $\lambda_k^j = 1$, as the indicator $\mathbb{1}_{g^j(x^j)\le 0} = 1$. Additionally, for all $k<j$, the solution $\lambda_k^k = 1$ is a valid solution. Restricting the solutions of $\lambda_1^1,\ldots,\lambda_K^K$ to this form forces the solution to lie on a hyperplane. We can enforce this constraint, which allows the variables $\lambda_1^1,\ldots,\lambda_K^K$ to be eliminated and the maximization to be expressed:

$$\max_{k\in\{1,\ldots,K\}}\left(\left(\sum_{j=k+1}^{K}\pi^k\right)\mathbb{1}_{g^k(\mathbf{x}^k)>0} + \sum_{j=1}^{k-1}\pi^j \mathbb{1}_{g^j(\mathbf{x}^j)\le 0}\right).$$

Substituting into

$$R(f^1,\ldots,f^K,x,y) = 1+\alpha\sum_{k=1}^{K} c_k - \sum_{k=1}^{K}\pi^k + \sum_{k=1}^{K}\pi^k \max\left(\mathbb{1}_{g^k(x^k)>0}, \mathbb{1}_{g^1(x^1)\le 0},\ldots,\mathbb{1}_{g^{k-1}(x^{k-1})\le 0}\right)$$

produces the empirical risk as shown in (6).

Details of Proposition 4.1 The linear program in (8) introduces new variables to replace the maximization functions in the reformulated empirical risk and the hinge losses. In particular, the variable γ_i replaces the maximization $\max_{k\in\{1,\ldots,K\}}\ldots$, the variables β_i^k captures the hinge-loss maximization $\max(1-g^k(\mathbf{x}_i^k),0)\ge \mathbb{1}_{g^k(\mathbf{x}_i^k)\le 0}$, and the variables κ_i^k captures the hinge-loss maximization $\max(1+g^k(\mathbf{x}_i^k),0)\ge \mathbb{1}_{g^k(\mathbf{x}_i^k)\ge 0}$. Introducing the variables $\gamma_i,\beta_i^k,\kappa_i^k$ allow the maximizations to be replaced with constraints. Additionally, we drop the constant terms in the optimization as these do not affect the functions $g^1,\ldots,g^K$ found by solving the optimization.