Exercise 1 (Support Vector Machine (SVM)) Minimization of convex functions : Karush-Kuhn-Tucker sufficient conditions

Let $f, -g_1, \ldots, -g_n$ be $\mathcal{C}^1$ convex functions and define the Lagrangian

$$\mathcal{L} : (x, \lambda) \mapsto f(x) - \sum_{i=1}^{n} \lambda_i g_i(x).$$

For any (x, λ), the Karush-Kuhn-Tucker conditions read :

1. $\forall i \in [\![1, n]\!] : g_i(x) \geqslant 0$;
2. $\forall i \in [\![1, n]\!] : \lambda_i \geqslant 0$;
3. $\nabla_x L(x, \lambda) = 0$;
4. $\min(\lambda_i, g_i(x)) = 0$ for $i = 1, \ldots, n$.

We know that, under the previous assumptions, KKT conditions are sufficient : if a couple $(\hat{x}, \hat{\lambda})$ fulfills the KKT conditions, then

$$\hat{x} \in \underset{\forall i \in [\![1,n]\!]: g_i(x) \geqslant 0}{\operatorname{argmin}} f(x) \quad \text{and} \quad \hat{\lambda} \in \underset{\lambda \geqslant 0}{\operatorname{argmax}} \inf_x \mathcal{L}(x, \lambda) \ .$$

Also, still under the previous assumptions, weak duality holds :

$$\sup_{\lambda \geqslant 0} \inf_x \mathcal{L}(x, \lambda) \leqslant \inf_x \sup_{\lambda \geqslant 0} \mathcal{L}(x, \lambda) = \inf_{\forall i \in [\![1,n]\!]: g_i(x) \geqslant 0} f(x).$$

Strong duality (i.e. equality holds) under additional assumptions.

Strong duality : If there exists a x such that $g_i(x) > 0$ for all $i \in \{1, \ldots, n\}$, then the KKT conditions are also necessary (i.e. $\hat{\lambda}$ exists and KKT conditions are satisfied by $(\hat{x}, \hat{\lambda})$) and

$$\sup_{\lambda \geqslant 0} \inf_x \mathcal{L}(x, \lambda) = \inf_x \sup_{\lambda \geqslant 0} \mathcal{L}(x, \lambda) \ .$$

Application to SVM

For any $w \in \mathbb{R}^p$, define the linear function $f_w(x) = \langle w, x \rangle$ from $\mathbb{R}^p$ to $\mathbb{R}$. For a given $R > 0$, we consider the set of linear functions $\mathcal{F} = \{f_w : \|w\| \leqslant R\}$. The aim of this exercise is to investigate the classifier $\widehat{h}_{\varphi,\mathcal{F}}(x) = \mathsf{sign}(\widehat{f}_{\varphi,\mathcal{F}}(x))$ where $\widehat{f}_{\varphi,\mathcal{F}}$ is solution to the convex optimisation problem

$$\widehat{f}_{\varphi,\mathcal{F}} \in \operatorname{argmin}_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \varphi(-y_i f(x_i)) \ ,$$

with $\varphi(x) = (1 + x)_+$ the *hinge* loss.

1. From the strong duality, prove that there exists $\lambda \geq 0$ such that
$$\widehat{f}_{\varphi,\mathcal{F}} \in \operatorname{argmin}_{f_w} \left\{ \frac{1}{n} \sum_{i=1}^{n} (1 - y_i f_w(x_i))_+ + \lambda \|w\|^2 \right\}.$$
2. Prove that $\widehat{f}_{\varphi,\mathcal{F}} = f_{\widehat{w}}$ where $\widehat{w}$ belongs to $V = \mathsf{Span}\{x_i : i = 1, \ldots, n\}$.
3. Prove that $\widehat{w} = \sum_{j=1}^{n} \widehat{\beta}_j x_j$ where $\widehat{\beta} = [\widehat{\beta}_1, \ldots, \widehat{\beta}_n]^\top$ is solution to
$$\widehat{\beta} = \operatorname{argmin}_{\beta \in \mathbb{R}^n} \left\{ \frac{1}{n} \sum_{i=1}^{n} (1 - y_i (K\beta)_i)_+ + \lambda \beta^\top K \beta \right\} \ ,$$
with K the Gram matrix $K = [\langle x_i, x_j \rangle]_{1 \leq i,j \leq n}$.

4. Check that this minimization problem is equivalent to

$$\widehat{\beta} = \operatorname*{argmin}_{\substack{\beta,\,\xi\,\in\,\mathbb{R}^n \text{ such that} \\ y_i(K\beta)_i \geqslant 1-\xi_i \\ \xi_i \geqslant 0}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \xi_i + \lambda \beta^\top K \beta \right\}.$$

5. Let us assume that K is not singular. From the KKT conditions, check that $\widehat{\beta}_i = y_i \widehat{\alpha}_i/(2\lambda)$, for $i = 1, \ldots, n$ with $\widehat{\alpha}_i$ fulfilling $\min(\widehat{\alpha}_i, y_i(K\widehat{\beta})_i - (1-\widehat{\xi}_i)) = 0$ et $\min(1/n - \widehat{\alpha}_i, \widehat{\xi}_i) = 0$.
6. Prove the following properties
   — if $y_i \widehat{f}_{\varphi,\mathcal{F}}(x_i) > 1$ then $\widehat{\beta}_i = 0$;
   — if $y_i \widehat{f}_{\varphi,\mathcal{F}}(x_i) < 1$ then $\widehat{\beta}_i = y_i/(2\lambda n)$;
   — in any case (in particular if $y_i \widehat{f}_{\varphi,\mathcal{F}}(x_i) = 1$), $0 \leqslant \widehat{\beta}_i y_i \leqslant 1/(2\lambda n)$.
7. From the strong duality, prove that $\widehat{\alpha}_i$ is solution to the dual problem

$$\widehat{\alpha} = \operatorname*{argmax}_{0\leqslant\alpha_i\leqslant 1/n} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{4\lambda} \sum_{i,j=1}^{n} K_{i,j} y_i y_j \alpha_i \alpha_j \right\}.$$