Convex Optimization
Exercises

Luis Antonio Ortega Andrés
Antonio Coín Castro

March 25, 2021

Luis Antonio: 50%, Antonio: 50%

Exercise 1. *Show that if S is an open set, its complement S^c is closed, and viceversa.*

Solution. Suppose S is an open set. We will prove that S^c is closed by showing that it coincides with its closure $\mathrm{cl}(S^c)$. On the one hand, we have by definition that $S^c \subset \mathrm{cl}(S^c)$, since

$$\mathrm{cl}(S^c) = \{x : \forall\delta\, B(x,\delta) \cap S^c \neq \emptyset\}.$$

On the other hand, let $x \in \mathrm{cl}(S^c)$ and $\delta > 0$. Then, we have:

$$B(x,\delta) \cap S^c \neq \emptyset \implies B(x,\delta) \not\subset S. \tag{1}$$

But since S is open, it holds that $\forall x \in S\ \exists\delta > 0$ such that $B(x,\delta) \subset S$. Therefore, as δ was arbitrary in (1), necessarily $x \notin S$, and so $x \in S^c$. As x was also arbitrary, we have shown that $cl(S^c) \subset S^c$, and hence that the complement of an open set is closed, as desired. The converse is obvious by taking complements. □

Luis Antonio: 50%, Antonio: 50%

Exercise 2. *If S_1, S_2 are convex subsets, prove that $S_1 \cap S_2$, $S_1 + S_2$ and $S_1 - S_2$ are also convex sets.*

Solution. We will begin with $S_1 \cap S_2$. Let $x, x' \in S_1 \cap S_2$ and $\lambda \in [0,1]$. By definition of $S_1 \cap S_2$ we have $x, x' \in S_1$ and $x, x' \in S_2$, and by convexity it holds that

$$\lambda x + (1-\lambda)x' \in S_1 \quad \text{and} \quad \lambda x + (1-\lambda)x' \in S_2$$

Therefore, $\lambda x + (1-\lambda)x' \in S_1 \cap S_2$, proving that the set is indeed convex.

Consider now $x, x' \in S_1 + S_2$. We can decompose them as $x = x_1 + x_2$ and $x' = x_1' + x_2'$, where $x_1, x_1' \in S_1$ and $x_2, x_2' \in S_2$. If $\lambda \in [0,1]$, using the convexity of S_1 and S_2 we have:

$$\begin{aligned}\lambda x + (1-\lambda)x' &= \lambda x_1 + \lambda x_2 + x_1' + x_2' - \lambda x_1' - \lambda x_2' \\ &= \underbrace{\lambda x_1 + (1-\lambda)x_1'}_{\in S_1} + \underbrace{\lambda x_2 + (1-\lambda)x_2'}_{\in S_2} \in S_1 + S_2.\end{aligned}$$

Lastly, the set $S_1 - S_2$ is shown to be convex using the same reasoning as we did for $S_1 + S_2$. $\square$

Luis Antonio: 50%, Antonio: 50%
Exercise 3. *If $f : S \to \mathbb{R}$ is a convex function on the convex set S, the set $\{x : \ x \text{ is a minimum of } f\}$ is a convex set.*

Solution. Let $M(f) = \{x \in S : \ x \text{ is a minimum of f}\}$ and suppose it is non-empty (otherwise there is nothing to prove, since the empty set is convex). Consider $x, x' \in M(f)$ and $\lambda \in [0, 1]$. Now, let $c \in \mathbb{R}$ be the minimal value attained by f, so that $f(x) = f(x') = c$. Using this fact and the convexity of f, we have:

$$f(\lambda x + (1-\lambda)x') \leq \lambda f(x) + (1-\lambda) f(x') = \lambda c + (1-\lambda)c = c,$$

but since c is the minimal value of f, we also have $f(\lambda x + (1-\lambda)x') \geq c$, so necessarily $f(\lambda x + (1-\lambda)x') = c$, and hence $\lambda x + (1-\lambda)x' \in M(f)$. As a result, $M(f)$ is a convex subset of S. $\square$

Luis Antonio: 50%, Antonio: 50%
Exercise 4. *Let $f : S \subset \mathbb{R}^d \to \mathbb{R}$ be a convex function on the convex set S, and suppose we extend it to $\hat{f} : \mathbb{R}^d \to \mathbb{R}$ as*

$$\hat{f}(x) = \begin{cases} f(x), & \text{if } x \in S, \\ +\infty, & \text{if } x \notin S. \end{cases}$$

Show that $\hat{f}$ is a convex function on $\mathbb{R}^d$.

Solution. Let $x, x' \in \mathbb{R}^d$ and $\lambda \in [0, 1]$. There are essentially two possibilities:

- If $x, x' \in S$, then $\lambda x + (1-\lambda)x' \in S$ by convexity, and looking at the definition of $\hat{f}$ and using the convexity of f, we have:

$$\begin{aligned} \hat{f}(\lambda x + (1-\lambda)x') &= f(\lambda x + (1-\lambda)x') \\ &\leq \lambda f(x) + (1-\lambda) f(x') \\ &= \lambda \hat{f}(x) + (1-\lambda)\hat{f}(x'). \end{aligned}$$

- If, say, $x \notin S$, then $\lambda \hat{f}(x) + (1-\lambda)\hat{f}(x') = +\infty$, and the condition for the convexity of $\hat{f}$ is trivially verified, since $\hat{f}(z) \leq +\infty$ for all z. The same happens if $x' \notin S$.

In any case, the line segment joining any two points on the graph of $\hat{f}$ lies above the graph between those two points, so $\hat{f}$ is convex. $\square$

Luis Antonio: 50%, Antonio: 50%
Exercise 5. *If Q is a symmetric, positive definite $d \times d$ matrix, show that $f(x) = x^T Q x, \ x \in \mathbb{R}^d$, is a convex function.*

Let $x, x' \in \mathbb{R}^d$ and $\lambda \in [0,1]$. We want to show that

$$f(\lambda x + (1-\lambda)x') \leq \lambda f(x) + (1-\lambda) f(x'),$$

or equivalently, that

$$f(\lambda x + (1-\lambda)x') - \lambda f(x) + (1-\lambda) f(x') \leq 0.$$

The result follows after some elementary algebraic manipulations:

$$f(\lambda x + (1-\lambda)x') - \lambda f(x) - (1-\lambda) f(x') =$$

$$= \lambda^2 x^T Q x + 2\lambda(1-\lambda)(x')^T Q x + (1-\lambda)^2 (x')^T Q x' - \lambda x^T Q x - (1-\lambda)(x')^T Q x' =$$
$$= \lambda(\lambda-1)\left[x^T Q x + (x')^T Q x' - 2x^T Q x'\right] = \underbrace{\lambda(\lambda-1)}_{\leq 0}\underbrace{(x-x')^T Q (x-x')}_{>0} \leq 0,$$

where in the last inequality we have used that Q is positive definite. Throughout the calculations we have also used that, since Q is symmetric, $(x')^T Q x = x^T Q x'$. □

Luis Antonio: 50%, Antonio: 50%

Exercise 6. *Given a quadratic form $q(w) = w^T Q w + bw + c$, with Q a symmetric $d \times d$ matrix, w, b being $d \times 1$ vectors and c a real number, derive its gradient and Hessian.*

Solution. Starting from the expanded form

$$q(w) = \sum_{i=1}^{d}\sum_{j=1}^{d} Q_{ij} w_i w_j + \sum_{i=1}^{d} b_i w_i + c$$

and taking into account that $Q_{ij} = Q_{ji}$, we can easily compute the partial derivatives

$$\frac{\partial q(w)}{\partial w_i} = 2Q_{ii} w_i + \sum_{\substack{j=1 \\ j \neq i}}^{d} w_j (Q_{ij} + Q_{ji}) + b_i = \sum_{j=1}^{d} 2Q_{ij} w_j + b_i, \quad i = 1, \ldots, d,$$

and

$$\frac{\partial^2 q(w)}{\partial w_i \partial w_j} = 2Q_{ij}, \quad i, j = 1, \ldots, d.$$

As a result, we have

$$\nabla q(w) = \left(\frac{\partial q(w)}{\partial w_1}, \ldots, \frac{\partial q(w)}{\partial w_d}\right)^T = 2\left(\sum_{j=1}^{d} Q_{1j} w_j, \ldots, \sum_{j=1}^{d} Q_{dj} w_j\right)^T + (b_1, \ldots, b_d)^T = 2Qw + b,$$

and

$$Hq(w) = \begin{pmatrix} \dfrac{\partial^2 q(w)}{\partial w_1 \partial w_1} & \cdots & \dfrac{\partial^2 q(w)}{\partial w_1 \partial w_d} \\ \vdots & & \vdots \\ \dfrac{\partial^2 q(w)}{\partial w_d \partial w_1} & \cdots & \dfrac{\partial^2 q(w)}{\partial w_d \partial w_d} \end{pmatrix} = \begin{pmatrix} 2Q_{11} & \cdots & 2Q_{1d} \\ \vdots & & \vdots \\ 2Q_{d1} & \cdots & 2Q_{dd} \end{pmatrix} = 2Q.$$

□

Luis Antonio: 50%, Antonio: 50%
Exercise 7. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a function and assume that $epi(f) \subset \mathbb{R}^d \times \mathbb{R}$ is convex. Prove that then f is convex.*

Solution. Recall that in this case

$$\operatorname{epi}(f) = \{(x,t) \in \mathbb{R}^d \times \mathbb{R} : \ t \geq f(x)\}$$

is the graph above f. Let $x, x' \in \mathbb{R}^d$ and consider $a = (x, f(x))$ and $b = (x', f(x'))$, observing that both of them are in $\operatorname{epi}(f)$. Then, since $\operatorname{epi}(f)$ is convex, for $\lambda \in [0,1]$ we have that $\lambda a + (1-\lambda) b \in \operatorname{epi}(f)$, which in turn implies that

$$(\lambda x + (1-\lambda)x', \ \lambda f(x) + (1-\lambda) f(x')) \in \operatorname{epi}(f).$$

Therefore, by definition of the epigraph, $f(\lambda x + (1-\lambda)x') \leq \lambda f(x) + (1-\lambda) f(x')$, and hence f is convex. □

Luis Antonio: 50%, Antonio: 50%
Exercise 8. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Prove that* $\operatorname{epi}(f)$ *is a closed set and that* $(x, f(x)) \in \partial \operatorname{epi}(f)$.

Solution. Let us show first that $\operatorname{epi}(f)$ is closed. There are two different approaches:

- We could use the well-known fact that convex functions on open sets are continuous, and the characterization of closed sets as sets that contain all the limiting points of sequences in them. Indeed, if $\{(x_n, t_n)\} \subset \operatorname{epi}(f)$ with $(x_n, t_n) \to (x,t)$, we have:

  $$f(x_n) \leq t_n \implies f(x) = \lim_{n\to\infty} f(x_n) \leq \lim_{n\to\infty} t_n = t \implies (x,t) \in \operatorname{epi}(f),$$

  where the limit can be taken inside the function argument by continuity.

- Alternatively, we could show that $\operatorname{epi}(f) = \operatorname{cl}(\operatorname{epi}(f))$. To prove the non-trivial inclusion, let $(x,t) \in \operatorname{cl}(\operatorname{epi}(f))$, which means that

  $$\operatorname{epi}(f) \cap B((x,t),\delta) \neq \emptyset \ \ \forall \delta > 0. \tag{2}$$

  Suppose by contradiction that $(x,t) \notin \operatorname{epi}(f)$ and observe that since f is convex, $\operatorname{epi}(f)$ is also convex (the proof is straightforward and similar to that of Exercise

7). Therefore, by the Projection Theorem[1] a minimum distance $\alpha > 0$ from (x, t) to $\mathrm{epi}(f)$ is attained, and so it holds that

$$\mathrm{epi}(f) \cap B((x, t), \alpha/2) = \emptyset,$$

contradicting (2).

Now, it is obvious that $(x, f(x)) \in \mathrm{epi}(f) = \mathrm{cl}(\mathrm{epi}(f))$, since $f(x) \geq f(x)$. To show that $(x, f(x)) \in \partial\,\mathrm{epi}(f)$, it suffices to show that it doesn't belong to $\mathrm{int}(\mathrm{epi}(f))$, since

$$\partial\,\mathrm{epi}(f) = \mathrm{cl}(\mathrm{epi}(f)) - \mathrm{int}(\mathrm{epi}(f)).$$

Indeed, if it weren't the case we could find a $\delta > 0$ such that

$$B((x, f(x)), \delta) \subset \mathrm{epi}(f),$$

so it would hold that, for example, $(x, f(x) - \delta/2) \in \mathrm{epi}(f)$. But by definition of the epigraph, this means that $f(x) \leq f(x) - \delta/2$, which is a contradiction given that $\delta > 0$. □

Luis Antonio: 50%, Antonio: 50%

Exercise 9. *Prove that if f is strictly convex, it has a unique global minimum.*

Solution. We will prove that for strictly convex functions defined on convex sets, there can be at most one global minimum. Note that it is still possible for a strictly convex function to have no minima (consider the function $x \mapsto e^x$ on $\mathbb{R}$).

Let $S \subset \mathbb{R}^d$ be convex and $f : S \to \mathbb{R}$ be strictly convex. We argue by contradiction and suppose $x^*, \tilde{x} \in S$ are two *distinct* global minimums of f, that is,

$$x^* \neq \tilde{x} \quad \text{and} \quad f(x^*) = f(\tilde{x}) \leq f(x) \quad \forall x \in S.$$

Now, since f is strictly convex, for any $\lambda \in (0, 1)$ we have

$$f(\lambda x^* + (1 - \lambda)\tilde{x}) < \lambda f(x^*) + (1 - \lambda) f(\tilde{x}) = f(x^*),$$

which contradicts the fact that x^* is a global minimum of f. Therefore, f can have at most one global minimum. □

Luis Antonio: 50%, Antonio: 50%

Exercise 10. *Let $f, g : S \subset \mathbb{R}^d \to \mathbb{R}$ be two convex functions on the convex set S . Prove that, as subsets, $\partial(f + g)(x) \subset \partial f(x) + \partial g(x)$ for any $x \in S$.*

Solution. This is (part of) the statement of the Moreau-Rockafellar Theorem that we saw in class. This is the hard part, since the other inclusion can be easily proved, and we do so here for completeness.

[1] We only need the existence part, which follows directly from Weierstrass' Theorem applied to the continuous function $z \mapsto \|z - x\|$ on the compact set $\{x' \in \mathrm{cl}(\mathrm{epi}(f)) : \|x' - x\| \leq \|z - x\|\}$ for a fixed $z \in \mathrm{epi}(f)$.

Fix $x \in S$ and let $c \in \partial f(x) + \partial g(x)$. We can decompose it as $c = c_1 + c_2$ with $c_1 \in \partial f(x)$ and $c_2 \in \partial g(x)$. Then, by definition of subgradient and the distributive property of the scalar product, we have

$$\begin{aligned}\forall x' \in S \quad (f+g)(x') &= f(x') + g(x') \\ &\geq f(x) + g(x) + c_1(x'-x) + c_2(x'-x) \\ &= (f+g)(x) + c(x'-x),\end{aligned}$$

so that $c \in \partial(f+g)(x)$. The proof of the other inclusion can be consulted in the book *Convex Analysis* by Rockafellar himself[2], but it is somewhat technical and it utilizes previous results. We found an alternative proof in simpler terms in Prof. Tieyong's lecture notes for a course in Optimization Theory at the Chinese University of Hong Kong[3], which we briefly summarize below.

Fixing x and assuming $c \in \partial(f+g)(x)$, two auxiliary sets are defined, namely

$$\Lambda_f := \{(x'-x, t) : \ t > f(x') - f(x) - c(x'-x)\}$$

$$\Lambda_g := \{(x'-x, t) : \ -t \geq g(x') - g(x)\}.$$

They are easily shown to be disjoint, non-empty and convex, so the separation theorem applies and there exists a hyperplane defined by coefficients a and b that separates them. Next, b is shown to be negative via contradiction, and then the desired decomposition is defined as

$$c_2 = \frac{a}{b}, \quad c_1 = c - c_2.$$

Lastly, these coefficients are shown to verify $c_1 \in \partial f(x)$ and $c_2 \in \partial g(x)$.

Prior to consulting any references we tried to prove the theorem by ourselves, and came up with the following incomplete proof. We defined a weighted decomposition of the coefficient c as

$$c_1 = \frac{f(x') - f(x)}{f(x') - f(x) + g(x') - g(x)} c \quad \text{and} \quad c_2 = \frac{g(x') - g(x)}{f(x') - f(x) + g(x') - g(x)} c.$$

It is clear that $c_1 + c_2 = c$, and using the hypothesis $c(x'-x) \leq f(x') - f(x) + g(x') - g(x)$ we arrive at

$$c_1(x'-x) = \frac{f(x') - f(x)}{f(x') - f(x) + g(x') - g(x)} c(x'-x) \leq f(x') - f(x),$$

and the analogous holds for c_2. The main problem here is that this decomposition is only valid when $f(x') - f(x) + g(x') - g(x) \neq 0$, which at first glance doesn't necessarily hold for all x, x'.

---

[2] Theorem 23.8 in Rockafellar, R. T. (1997). *Convex analysis* (Vol. 36), p. 223. Princeton University Press.

[3] `https://www.math.cuhk.edu.hk/course_builder/1819/math4230/subgrad.pdf`

Luis Antonio: 50%, Antonio: 50%
Exercise 11. *Compute the proximal of $f(x) = 0$ and of $g(x) = \frac{1}{2}\|x\|^2$.*

Solution. Recall that the proximal operator can be expressed as

$$\operatorname{prox}_f(x) = \arg\min_z \left\{ f(z) + \frac{1}{2}\|z - x\|^2 \right\}.$$

The strategy will be to compute and minimize this expression in each case.

- On the one hand, we have $\operatorname{prox}_f(x) = \arg\min_z \left\{\frac{1}{2}\|z - x\|^2\right\}$, but $\|z - x\|^2$ is clearly minimized at $z = x$, so $\operatorname{prox}_f(x) = x$ for all x.
- On the other hand, $\operatorname{prox}_g(x) = \arg\min_z\{\frac{1}{2}\|z\|^2 + \frac{1}{2}\|z - x\|^2\}$. We know that for differentiable functions the minimizers obey the first order condition $\nabla = 0$, so:

  $$0 = \nabla_z \left(\frac{1}{2}\|z\|^2 + \frac{1}{2}\|z - x\|^2 \right) = z + z - x \implies z = \frac{1}{2}x.$$

  The point $x/2$ is indeed a minimizer of our function in z, for example because its Hessian matrix is $2I$, which is positive definite. Therefore, $\operatorname{prox}_g(x) = x/2$ for all x.

Luis Antonio: 50%, Antonio: 50%
Exercise 12. *Assume that f is convex. Prove that for any $\lambda > 0, \partial(\lambda f)(x) = \lambda\partial f(x)$ as subsets.*

Solution. Instead of showing the double inclusion, we prove the equivalence directly. We have:

$$\begin{aligned}
\partial(\lambda f)(x) &= \{c \in \mathbb{R}^d : \ \lambda f(x') \geq \lambda f(x) + c(x' - x) \ \ \forall x' \in S\} && \text{[Definition]} \\
&= \{c \in \mathbb{R}^d : \ f(x') \geq f(x) + \lambda^{-1} c(x' - x) \ \ \forall x' \in S\} && \text{[Divide by } \lambda > 0] \\
&= \{\lambda c \in \mathbb{R}^d : \ f(x') \geq f(x) + c(x' - x) \ \ \forall x' \in S\} && \text{[Rename]} \\
&= \lambda \partial f(x). && \text{[Definition]}
\end{aligned}$$

□

Luis Antonio: 50%, Antonio: 50%
Exercise 13. *Compute the proximals of the hinge $f(x) = max\{0, -x\}$ and the ϵ-insensitive $g(x) = max\{0, |x| - \epsilon\}$ loss functions.*

Solution. We proceed as in Exercise 11. Since both losses are convex and defined on (the open set) $\mathbb{R}$, we know that the proximal operator is well defined.

- On the one hand, we have $\operatorname{prox}_f(x) = \arg\min_z\{\max\{0, -z\} + \frac{1}{2}(z - x)^2\} = \arg\min_z h(z)$. We are looking for the minimizer of the function

  $$h(z) = \begin{cases} \frac{1}{2}(z - x)^2, & \text{if } z \geq 0, \\ -z + \frac{1}{2}(z - x)^2, & \text{if } z \leq 0. \end{cases}$$

If the minimizer is attained at $z > 0$, then clearly $z = x$, meaning that $\text{prox}_f(x) = x$ for $x > 0$. If it is attained at $z < 0$, we have:

$$0 = h'(z) = -1 + z - x \implies z = x + 1,$$

which implies that $\text{prox}_f(x) = x + 1$ for $x < -1$. The value of the proximal at the remaining values must necessarily be the only point of non-differentiability of h, namely, 0. Thus, we have shown that

$$\text{prox}_f(x) = \begin{cases} x+1, & \text{if } x \leq -1, \\ 0, & \text{if } -1 \leq x \leq 0, \\ x, & \text{if } x \geq 0. \end{cases}$$

- On the other hand, $\text{prox}_g(x) = \arg\min_z\{\max\{0, |z| - \epsilon\} + \frac{1}{2}(z-x)^2\} = \arg\min_z h(z)$. In this case, we seek the minimizer of

$$h(z) = \begin{cases} \frac{1}{2}(z-x)^2, & \text{if } |z| \leq \epsilon, \\ z - \epsilon + \frac{1}{2}(z-x)^2, & \text{if } z \geq \epsilon, \\ -z - \epsilon + \frac{1}{2}(z-x)^2, & \text{if } z \leq -\epsilon. \end{cases}$$

  Following the same reasoning as before, it is immediate to see that $\text{prox}_g(x) = x$ for $|x| < \epsilon$. If the minimizer is attained when $z > \epsilon$, we have $0 = h'(z) = 1 + z - x$, so $z = x - 1$, and $\text{prox}_g(x) = x - 1$ for $x > 1 + \epsilon$. Similarly, $\text{prox}_g(x) = x + 1$ for $x < -1 - \epsilon$.

  For the points of non-differentiability, observe that if $\epsilon < x < 1 + \epsilon$ the minimizer is attained at $z = \epsilon$ (the only possible value), while for $-1 - \epsilon < x < -\epsilon$ it is attained at $z = -\epsilon$. Thus, we have

$$\text{prox}_f(x) = \begin{cases} x+1, & \text{if } x \leq -1-\epsilon, \\ -\epsilon, & \text{if } -1-\epsilon \leq x \leq -\epsilon, \\ x, & \text{if } |x| \leq \epsilon, \\ \epsilon, & \text{if } \epsilon \leq x \leq 1+\epsilon, \\ x-1, & \text{if } x \geq 1+\epsilon. \end{cases}$$

Luis Antonio: 50%, Antonio: 50%

Exercise 14. *If $(p_1, \ldots, p_n)$ is a probability distribution, prove that its entropy $H(p_1, \ldots, p_n) = -\sum_{i=1}^n p_i \log p_i$ is a concave function. Show also that its maximum is $\log n$, attained when $p_i = \frac{1}{n}$ for all i.*

Solution. In the first place, we know that H is twice-differentiable on the open set $(0,1)^n$. Moreover, since

$$\frac{\partial H}{\partial p_i} = -\log p_i - 1 \quad \text{and} \quad \frac{\partial^2 H}{\partial p_i \partial p_j} = \frac{-1}{p_i}\delta_{ij},$$

we have

$$\text{Hess}(H) = \text{diag}\left\{\frac{-1}{p_i}\right\}_{i=1,\dots,n} \prec 0.$$

Since the Hessian is negative definite, H is concave. To solve the constrained optimization problem, we consider the corresponding Lagrangian

$$L(p_1, \dots, p_n; \mu) = -\sum_{i=1}^{n} p_i \log p_i + \mu\left(\sum_{i=1}^{n} p_i - 1\right),$$

where the restriction $\sum p_i = 1$ appears because the p_i are a probability distribution. Now, we have

$$\frac{\partial L}{\partial p_i} = -\log p_i - 1 + \mu, \quad i = 1, \dots, n \quad \text{and} \quad \frac{\partial L}{\partial \mu} = \sum_{i=1}^{n} p_i - 1.$$

To find the point where all derivatives vanish simultaneously, we have to solve the system

$$\begin{cases} \log p_i & = \mu - 1, \quad i = 1, \dots, n, \\ \sum p_i & = 1. \end{cases}$$

Looking at the first equation, since the right hand side doesn't depend on i, it holds that $p_i = p_j = e^{\mu-1} \geq 0$ for all i, j, that is, all the p_i are equal. But since they must add up to 1, the only possibility is that $p_i = 1/n$ for all i. The restriction $p_i \geq 0$ is clearly satisfied, and since H is concave the resulting value is indeed a maximum, whose value is

$$H(1/n, \dots, 1/n) = -\frac{1}{n}\sum_{i=1}^{n} \log\frac{1}{n} = \frac{n \log n}{n} = \log n.$$

$\square$

Luis Antonio: 50%, Antonio: 50%

Exercise 15. *Work out the dual problem for the hard SVC problem*

$$\min_{w,b}\left\{\frac{1}{2}\|w\|^2\right\}$$

subject to $y_p(w^T x_p + b) \geq 1$. *What are the KKT conditions?*

Solution. In this case, the Lagrangian is

$$L(w, b; \alpha) = \frac{1}{2}\|w\|^2 + \sum_p \alpha_p(1 - y_p(w^T x_p + b)),$$

with $\alpha_p \geq 0$. To get the dual function we solve $\nabla_w L = 0$ and $\frac{\partial L}{\partial b} = 0$, which yields:

$$0 = \nabla_w L = w - \sum_p \alpha_p y_p x_p \implies w = \sum_p \alpha_p y_p x_p,$$
$$0 = \frac{\partial L}{\partial b} = \sum_p \alpha_p y_p.$$

Substituting both expressions back in the Lagrangian, we arrive at the dual function

$$\begin{aligned}
\Theta(\alpha) &= \frac{1}{2}\sum_{p,q} \alpha_p \alpha_q y_p y_q x_p^T x_q + \sum_p \alpha_p - \sum_{p,q} \alpha_p \alpha_q y_p y_q x_p^T x_q + b \overbrace{\sum_p \alpha_p y_p}^{0} \\
&= -\frac{1}{2}\sum_{p,q} \alpha_p \alpha_q y_p y_q x_p^T x_q + \sum_p \alpha_p \\
&= -\frac{1}{2}\alpha^T Q \alpha + \alpha^T \mathbf{1},
\end{aligned}$$

subject to $\sum_p \alpha_p y_p = 0$ and $\alpha_p \geq 0$ for all p. In the above expression, Q is the symmetric matrix given by $Q_{pq} = y_p y_q x_p^T x_q$, $\alpha = (\alpha_1, \ldots, \alpha_p)^T$ and $\mathbf{1} = (1, \ldots, 1)^T$. Flipping the sign to get a minimization problem, the dual problem has the following expression:

$$\min_{\alpha \in \mathbb{R}^N} \left\{ \frac{1}{2}\alpha^T Q \alpha - \alpha^T \mathbf{1} \right\} \quad \text{subject to} \quad \begin{cases} \sum_p \alpha_p y_p = 0, \\ \alpha_p \geq 0, \;\; p = 1, \ldots, N. \end{cases}$$

The KKT conditions in this case are:

$$\begin{cases} w^* = \sum_p \alpha_p^* y_p x_p, \\ \alpha_p^* (1 - y_p((w^*)^T x_p + b^*)) = 0, \;\; p = 1, \ldots, N. \end{cases}$$

The first one allows us to recover the primal solution from the dual one, and the rest are useful to define the support vectors: those points x_p for which $\alpha_p^* > 0$ must necessarily be in one of the supporting hyperplanes $(w^*)^T x_p + b^* = \pm 1$ (assuming labels $y_p = \pm 1$). Looking at the first condition again, we can also say that the points for which $\alpha_p^* = 0$ have no impact on the model. □