Ejercicios de Optimización Convexa y Restringida

1 S abierto $\longleftrightarrow$ S^c cerrado

Luis Castillo y Adrián Rubio

Entrega

Ejercicios

Optimización

28/3/2022

ϕ

$\longrightarrow x \in S^c$ ✓

hay $\delta > 0$

$(x, \delta) \cap S^c = \phi \longrightarrow$

$\longrightarrow$ aplicando def. convexidad

$\in S_2$

$S_1 \cap S_2$ convexo ✓

$$\underbrace{\lambda x + (1-\lambda) x'}_{x, x' \in S_1} + \underbrace{\lambda z + (1-\lambda) z'}_{z, z' \in S_2} \in S_1 + S_2$$

✓

$\xrightarrow{\text{continúa.}}$

Ejercicios de Optimización Convexa y Restringida

1 S abierto $\longleftrightarrow$ S^c cerrado

① S abierto $\longrightarrow S^c \supset cl(S^c)$

$x \in cl(S^c) \longrightarrow \forall \delta > 0 \ \boxed{B(x,\delta) \cap S^c \neq \phi}$

$\longrightarrow$ para ningún δ $B(x,\delta) \subset S \longrightarrow x \in S^c$ ✓

② $S^c = cl(S^c) \longrightarrow S$ abierto

$x \in S \longrightarrow x \notin S^c \longrightarrow x \notin cl(S^c) \longrightarrow$ no hay $\delta > 0$

$B(x,\delta) \cap S^c = \phi \longrightarrow$

$\longrightarrow B(x,\delta) \subset S \longrightarrow S$ abierto ✓

2 S_1, S_2 convexos

2.1 $S_1 \cap S_2$ convexo

$x, x' \in S_1 \cap S_2 \longrightarrow \underbrace{x, x' \in S_1}_{\lambda x + (1-\lambda)x' \in S_1}$ y $\underbrace{x, x' \in S_2}_{\lambda x + (1-\lambda)x' \in S_2}$ $\xrightarrow{\text{aplicando def. convexidad}}$

$0 \leq \lambda \leq 1$

$\lambda x + (1-\lambda)x' \in S_1 \cap S_2$ convexo ✓

2.2 $S_1 + S_2$ convexo

$S_1 + S_2 = \{x + z \,;\, x \in S_1, z \in S_2\}$

$x + z \in S_1 + S_2$

$x' + z' \in S_1 + S_2$

$0 \leq \lambda \leq 1$

$\lambda(x+z) + (1-\lambda)(x'+z') \in S_1 + S_2$

$\lambda x + \lambda z + (1-\lambda)x' + (1-\lambda)z' \in S_1 + S_2$

$\underbrace{\lambda x + (1-\lambda)x'}_{x, x' \in S_1} + \underbrace{\lambda z + (1-\lambda)z'}_{z, z' \in S_2} \in S_1 + S_2$

✓

* continúa. $\longrightarrow$

2.3 $S_1 - S_2$ Convexo

$$S_1 - S_2 = \{x - z : x \in S_1,\ z \in S_2\}$$

$$x - z \in S_1 - S_2$$

$$x' - z' \in S_1 - S_2$$

$$\lambda(x - z) + (1-\lambda)(x' - z') \in S_1 - S_2$$

$0 \le \lambda \le 1$

$$\underbrace{\lambda x + (1-\lambda)x'}_{x'' \in S_1} - \underbrace{[\lambda z + (1-\lambda)z']}_{z'' \in S_2} \in S_1 - S_2$$

$$= x'' - z'' \in S_1 - S_2 \quad \checkmark$$

3

S convexo $\longrightarrow$ Suponemos S no vacío $\boxed{S \text{ nEC}}$

$f: S \to \mathbb{R}$ convexa

$A = \{x : x \text{ es mínimo de } f\}$ ¿A convexo?

- Cogemos $\boxed{x, x' \in A}$ y aplicando <u>definición de f convexa</u>: $(0 \le \lambda \le 1)$

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

- Por unicidad del valor mínimo $m = \min_{x \in S} f(x)$, tenemos $\boxed{m = f(x) = f(x')}$:

$$\underbrace{f(\lambda x + (1-\lambda)x') \le}\ \lambda m + (1-\lambda)m = \cancel{\lambda m} + m - \cancel{\lambda m} = \underline{m}$$

- Pero como m es por definición el mínimo, únicamente se puede dar el caso de igualdad:

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

- Como $x, x' \in A$, tenemos que $\boxed{\lambda x + (1-\lambda)x' \in A}$, es decir, la definición de <u>conjunto convexo</u>

$\boxed{A \text{ convexo}}$ $\checkmark$

4 $q(w) = w^t Q w + b \cdot w + c$ | Da el gradiente y el Hessiano.

Forma cuadrática.

$$q(w) = \sum_i \sum_j Q_{ij} w_i w_j = \underbrace{\sum_i Q_{ii} w_i^2}_{(1)} + \underbrace{\sum_i \sum_{\substack{j \\ i \neq j}} Q_{ij} w_i w_j}_{(2)}$$

$$\frac{\partial q}{\partial w_k} = \underbrace{2 Q_{kk} w_k}_{(1)} + \sum_{\substack{j \\ j \neq k}} Q_{kj} w_j + \sum_{\substack{i \\ i \neq k}} Q_{ik} w_i =$$

si y solo si $i = k$

- $$\frac{\partial q}{\partial w_k} = 2 Q_{kk} w_k + \sum_{\substack{j \\ j \neq k}} (Q_{kj} + Q_{jk}) w_j$$

Así que

$$\boxed{\nabla q(w) = \left(2 \sum_{j=1}^{d} Q_{1j} w_j + b_1, \dots, 2 \sum_{j=1}^{d} Q_{dj} w_j + b_d \right) = 2Qw + b}$$

- $$\frac{\partial q}{\partial w_i \partial w_j} = 2 Q_{ij} \quad \forall i, j \in [1, d]$$

Así que escribiendo en el Hessiano:

$$H(q(w)) = \begin{bmatrix} 2Q_{11} & \cdots & 2Q_{1d} \\ \vdots & & \\ 2Q_{d1} & \cdots & 2Q_{dd} \end{bmatrix}$$

5 $(p_1, \dots, p_k)$ distribución de prob. $\longrightarrow$ $\boxed{\sum_{i=1}^{k} p_i = 1}$ Restricción de igualdad

Entropía $\boxed{H(p_1, \dots, p_k) = -\sum_{i=1}^{k} p_i \log p_i}$

Demuestra que H es cóncava

$\boxed{\text{Una función } H \text{ es cóncava} \longleftrightarrow -H \text{ es convexa.}}$

Teorema

Si f tiene segunda derivada positiva en todo su dominio, entonces f es convexa

Denotemos $\boxed{f(x) = -x \log(x)}$

$$f'(x) = -x \log(e) \frac{1}{x} - \log(x) = -\log(x) - \log(e)$$

$$f''(x) = -\log(e) \frac{1}{x} < 0 \quad \forall x > 0$$

$H(x) = \sum_{x \in |X|} f(P_x(x))$, es decir, como el dominio de x es siempre positivo al aplicarlo a una distribución de probabilidad, tenemos que

$$H(P) = H(p_1, \dots, p_k) < 0, \text{ es decir, } \boxed{H(P) \text{ es cóncava } \forall p_i}$$

Prueba que su máximo es $\log k$, obtenido cuando $p_i = \frac{1}{k}$ $\forall i$

① Definimos un problema de optimización restringida

$$\min -H(P) \quad \text{s.t.} \quad \sum_i p_i = 1$$

② Construimos el lagrangiano

$$L(p_1, \dots, p_k; \lambda) = \sum_1^k (p_i \log p_i) + \lambda \left(1 - \sum_1^k p_i\right)$$

③ $\boxed{0 = \frac{\partial L(P)}{\partial p_i}}$

$$\frac{\partial L}{\partial p_i} = \log p_i + 1 - \lambda = 0 \quad \text{con } 1 \leq i \leq k$$

$$\log p_i = \lambda - 1$$

$$p_i = e^{\lambda - 1}$$

④ Imponemos la restricción $\sum_i^k p_i = 1$

$$\sum_{i=1}^{k} p_i = e^{\lambda - 1} \cdot k = 1$$

$$e^{\lambda - 1} = \boxed{\frac{1}{k} = p_i}$$

Para que $-H$ alcance su mínimo

Para que H alcance su máximo

$$\boxed{H\left(p_1 = \frac{1}{k}, \dots, p_k = \frac{1}{k}\right)} = -\cancel{k} \cdot \frac{1}{\cancel{k}} \cdot \log \frac{1}{k} = -[\underbrace{\cancel{\log 1}}_{0} - \log k] = \boxed{\log k}$$

$\boxed{\text{El máximo de } H \text{ es } \log k}$

6

$$\min f(x,y) = x^2 + 2y^2 + 4xy$$
$$\text{s.t. } x+y=1$$
$$x \geq 0,\ y \geq 0$$

→ Reescribo inecuaciones tal que $\boxed{\begin{array}{l} -x \leq 0 \\ -y \leq 0 \end{array}}$

f no convexa, 2 restricciones convexas de desigualdades y una ecuación afín. Hay puntos de Slater.

Resolvemos encontrando un punto KKT:

1. Lagrangiano

$$L(x,y,\alpha,\beta,\gamma) = x^2 + 2y^2 + 4xy - \alpha x - \beta y + \gamma(1-x-y)$$

2. Ecuaciones con gradiente de L

$$0 = \frac{dL}{dx} = 2x + 4y - \alpha - \gamma$$
$$0 = \frac{dL}{dy} = 4y + 4x - \beta - \gamma$$
$$0 = \frac{dL}{d\alpha} = \boxed{\alpha x = 0}$$
$$0 = \frac{dL}{d\beta} = \boxed{\beta y = 0}$$
$$0 = \frac{dL}{d\gamma} = 1 - x - y \rightarrow \boxed{1 = x + y}$$

$\boxed{\alpha, \beta \geq 0}$

3. Consideramos los siguientes casos para los multiplicadores de inecualdad (y ver los puntos KKT)

3

(1) $\boxed{\alpha = 0 = \beta}$ (Restamos 2 primeras ecuaciones) → $2x = 0$ → $\boxed{x=0}$ → 5ª ec. → $x+y=1$ → $\boxed{y=1}$ — Punto KKT $(0,1)$

(2) $\boxed{\alpha > 0,\ \beta > 0}$ $x = y = 0$ $\boxed{x+y=1}$ → No se satisface → No podemos sacar punto KKT

(3) $\boxed{\alpha > 0,\ \beta = 0}$ → $\alpha > 0$ → $\boxed{x=0}$ → $\cancel{x}^{0} + y = 1$ → $\boxed{y=1}$ — Punto KKT $(0,1)$

(4) $\boxed{\alpha = 0,\ \beta > 0}$ → $\beta > 0$ → $\boxed{y=0}$ → 5ª ec. $x + \cancel{y}^{0} = 1$ → $\boxed{x=1}$ — Punto KKT $1,0$

4. Determinamos cuál de los puntos KKT es el óptimo

$\min f(x,y)$
- → $(x,y) = (0,1)$ → $0 + 2 \cdot 1^2 + 0 = 2$
- → $(x,y) = (1,0)$ → $1^2 + 0 + 0 = \boxed{1}$ mínimo

El punto óptimo es $(1,0)$

SOLUCIÓN

7 Hard SVC Problem

$$1 \le p \le N \quad \min \frac{1}{2}\|w\|^2 \quad \text{s.t.} \quad y^p(w \cdot x^p + b) - 1 \ge 0$$

1 Lagrangiano

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum \underbrace{\alpha_p}_{\geq 0} \left[y^p(w \cdot \underbrace{x^p}_{\geq 0} + b) - 1 \right]$$

$$= \boxed{\frac{1}{2} w w - \sum \alpha_p \cdot y^p \cdot w \cdot x^p} + \underbrace{b \sum \alpha_p y_p}_{0} + \sum \alpha_p$$

2 Gradiente

$$\nabla_w L = w - \sum \alpha_p y^p x^p = 0 \longrightarrow \boxed{w = \sum \alpha_p y^p x^p}$$

$$\frac{\partial L}{\partial b} = -\sum \alpha_p y^p = 0 \quad \text{Restricción}$$

3 Enchufamos 2 en 1

$$\rightarrow w\left(\frac{1}{2} w - \sum \alpha_p y^p \cdot x^p\right) - \frac{1}{2} w w$$

$$\rightarrow L(w,b,\alpha) = \sum \alpha_p - \frac{1}{2}\left(\sum_p \alpha_p y^p x^p\right)\left(\sum_q \alpha_q y^q x^q\right) =$$

$$= \sum \alpha_p - \frac{1}{2} \sum_{p,q} \alpha_p y^p \alpha_q y^q x^p x^q = \boxed{-\alpha}$$

$$\max_{\alpha} \underbrace{\sum \alpha^p - \frac{1}{2} \sum \alpha_p \alpha_q y^p y^q x^p x^q}_{\Theta(\alpha):\ \text{Función dual}} \quad \text{s.t.} \quad \alpha^p \geq 0, \quad \sum \alpha_p y^p = 0, \quad 1 \le p \le N$$

$$L(w, \lambda, \mu) = \sum \alpha^p - \frac{1}{2} \sum \alpha_p \alpha_q y^p y^q x^p x^q - \lambda \alpha^p + \mu \sum \alpha_p y^p$$

KKT
$$\begin{cases} w^* = \sum_p \alpha_p y_p x_p \\ \alpha_p = -y_p(w^* x_p + b^*) + 1 \end{cases}$$

8

$$\min \; c\cdot x \quad \text{s.t.} \quad \underbrace{x \geq 0}_{\lambda}, \; \underbrace{Ax \leq b}_{\mu} \qquad \text{con } x \in \mathbb{R}^d;\ A\ m\times d;\ b \in \mathbb{R}^m$$

multiplicadores: λ, μ

1 Escribo Lagrangiano

$$L(x,\lambda,\mu) = c\cdot x - \sum_{i=1}^{d} \lambda_i x_i + \sum_{j=1}^{m} \mu_j \underbrace{(a_j \cdot x - b_j)}_{\leq 0}$$

2 $\nabla L = 0$

$$\underbrace{\nabla_x L}_{d} = \underbrace{c}_{d} - \underbrace{\lambda}_{d} + \underbrace{A^t}_{d\times m} \cdot \underbrace{\mu}_{m\times 1} = 0$$

Dimensiones

Puedo escribirlo para cada elemento.

$$\frac{\partial}{dx_j}(\mu_j \cdot a_{ji} \cdot x_i - b_j) = (\mu_j a_{ji})$$

3 Reorganizo término del L

$$L(x,\lambda,\mu) = x\cdot c - x\cdot\lambda + x(A^t\cdot\mu) = x\cdot\cancel{\left(c - \lambda + A^t\cdot\mu\right)}^{\,0} - b\cdot\mu = -b\cdot\mu$$

$$\sum_{j=1}^{m} \mu_j \cdot x \cdot a_j = x \cdot \sum_{1}^{m} a_j \mu_j$$

dot product

$$\max \theta(\lambda,\mu) = -b\mu \quad \text{s.t.} \quad c = \lambda - A^t\cdot\mu$$

Problema dual.

Sin embargo aún contamos con λ en la restricción, y como los lambdas son positivos, puedo reescribir:

$$\max \theta(\mu) = -b\mu$$

que es igual a

$$\min \theta(\mu) = b\mu \quad \text{s.t.} \quad A^t\cdot\mu \leq -c$$

SOLUCIÓN

9) El SVC primal =>

$$\min_{w} \frac{1}{2}\|w\|^2 + C\sum_p \xi_p \quad \text{s.t. } y^p(wx^p+b) \geq 1-\xi_p, \quad \xi_p \geq 0$$

Usando:

$$C = \alpha_p^* + \beta_p^*$$

$$\xi^* = 1 - y^p(w^* \cdot x^p + b^*)$$

$$w^* = \sum_p \alpha_p^* y^p x^p$$

$$Q = \sum_{p,q} x^p x^q y^p y^q$$

Que usamos para transformar la ecuación del SVC primal:

$$\frac{1}{2}\|w\|^2 + C\sum_p \xi_p = \frac{1}{2}\sum_{pq} \alpha_p^* y^p x^p y^q x^q \cdot \alpha_q^{*T} \cdot \sum_p \left(-y^p(w^* x^p + b^*) + 1\right) =$$

$$= \frac{1}{2}\alpha^T Q \alpha + \alpha_p^* \sum_p \left(1 - y^p\left(\sum_q \alpha_q^* x^q y^q x^p + b^*\right)\right) + \beta_p \sum_p \left(1 - y^p(w^* \cdot x^p + b^*)\right) =$$

$$= \cdots = \frac{1}{2}\alpha^T Q \alpha + \sum_p \alpha_p^* - \alpha^T Q \alpha + \sum_p \beta_p^* \xi_p^* + b^* \sum_p \alpha_p^* y_p$$

$$= \sum_p \alpha_p^* - \frac{1}{2}\alpha^T Q \alpha + \cancel{\sum_p \beta_p^* \xi_p^*}^{\,0} + \cancel{b^* \sum_p \alpha_p^* y_p}^{\,0}$$

$$\theta(\alpha) = \sum_p \alpha_p^* - \frac{1}{2}\alpha^T Q \alpha$$

Puesto que minimizar un problema $\equiv$ maximizar el mismo problema $\cdot(-1)$

$$\min f = \max(-f)$$

Vemos que el problema primal y dual son equivalentes.

10 Ridge Regression con optimización con restricciones

$$\boxed{\min_{w} \frac{1}{2} mse(w) \quad s.t. \quad \|w\|^2 \leq \rho}$$

- Lagrangiano

$$L(w,\lambda) = \frac{1}{2} mse(w) + \frac{\lambda}{2}\left(\|w\|^2 - \rho\right)$$

$\longrightarrow$ ponemos 2 para simplificar luego con el gradiente.

- KKT's

$$0 = \nabla_w L(w,\lambda) = \frac{1}{N} x^t (x w^* - \underbrace{y}_{\text{Target}}) + \lambda^* w^* \quad = \star$$

$$0 = \lambda^* (\|w^*\| - \rho)$$

- $\boxed{\text{Si } \lambda^* > 0}$

$$\|w^*\|^2 = \rho \longrightarrow w^* \neq 0$$

$$\star \longrightarrow = \frac{1}{N}(x^t \cdot x + \lambda^* I) w^* - x^t \cdot y = 0 \rightarrow$$

$$\rightarrow w^* = \left[\frac{1}{N} \cdot x^t x + \cancel{\lambda^* I}\right]^{-1} \cdot x^t \cdot y$$

$$\rightarrow w^* \min \frac{1}{2N} \|x^t \cdot w - y\| + \lambda^* \|w\|^2$$

$\longrightarrow$ w^* es la solución Ridge para $\lambda^* > 0$

- $\boxed{\text{Si } \lambda^* = 0}$

$$\hookrightarrow \|w^*\|^2 \leq \rho$$

$$\star \longrightarrow = \frac{1}{N}(x^t \cdot x + I) w^* - x^t - y = 0$$

$$\longrightarrow w^* \min \frac{1}{2N} \|x^t w - y\|$$

$\longrightarrow$ w^* Solución de Least Squares

11

Q simetrica, +, d×d
matriz.

Teorema:

$$\boxed{f \text{ convex sii } f(z) - f(x) - \nabla f(x)\cdot(z-x) \geq 0}$$

$$z^tQz - x^tQx - 2x^tQ(z-x) = z^tQz - \cancel{x^tQx} - 2x^tQz + \cancel{2x^tQx} =$$

$$= (z-x)^t \cdot Q \cdot (z-x) \geq 0$$

Ya que Q es definida positiva

Y por tanto:

$$f(z) \geq f(x) + \nabla f(x)\cdot(z-x) \geq 0$$

Ejercicios de Optimización

Adrian Rubio y Luis Castillo

March 28, 2022

Luis: 50%, Adrian: 50%

Exercise 12. *Let $f : R^d \to R$ be a function and assume that epi$(f) \subset R^d \times R$ is convex. Prove that then f is convex.*

Solution. We know that in this case

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

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

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

By definition of the epigraph, $f(\lambda x + (1 - \lambda)x') \leq \lambda f(x) + (1 - \lambda)f(x')$, therefore f is convex.

Luis: 50%, Adrian: 50%

Exercise 13. *Let $f : R^d \to R$ be a convex function. Prove that* $\operatorname{epi}(f)$ *is a closed set and that* $(x, f(x)) \in \partial \operatorname{epi}(f)$.

Solution. First, lets show that $\operatorname{epi}(f)$ is closed:

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.

We know that $(x, f(x)) \in \operatorname{epi}(f) = \operatorname{cl}(\operatorname{epi}(f))$, since $f(x) \geq f(x)$. To show that $(x, f(x)) \in \partial \operatorname{epi}(f)$, it's enough proof that it doesn't belong to $\operatorname{int}(\operatorname{epi}(f))$, since

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

14

Demuestra que si f es estrictamente convexa $\longrightarrow$ único mínimo.

- Suponemos $\boxed{x \neq z}$ x y z son mínimos, $\lambda \in (0,1)$. Aplicando def. de función estri. convexa:

$$f(x\lambda + (1-\lambda)z) < \lambda f(x) + (1-\lambda) f(z) = f^*$$

Si el mínimo es único

$$f(x) = f(z) = f^*$$

(Ya que aunque x y z sean distintos, el valor de la función es único)

Por contradicción, sabemos que no puede ser menor que el mínimo, y solo puede haber un mínimo global.

Luis: 50%, Adrian: 50%

Exercise 15. *Let $f, g : S \subset R^d \to R$ 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 the Moreau-Rockafellar Theorem.

First, $x \in S$ 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:

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

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 shown to be disjoint, non-empty and convex, so the separation theorem applies and there exists an 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

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

16

- $\boxed{\text{prox}_0(x) = \text{argmin}_z\, 0 + \frac{1}{2}\|z-x\|^2 = x}$

Usando definición de proximal alternativa

- $\text{prox}_{\|\cdot\|^2}(x) = \text{argmin}_z \underbrace{\frac{1}{2}\|z\|^2 \cdot \frac{1}{2}\|z-x\|^2}_{g(z)}$

$0 = \nabla g(z) = z + (z-x) = 2z - x \; ; \; z = \frac{1}{2}x$

$$\boxed{\text{prox}_{\|\cdot\|^2} = \frac{1}{2}x}$$

17 Asume f convexa. Demuestra que $\forall\ \lambda > 0$, $\partial(\lambda f)(x) = \lambda \cdot \partial f(x)$.

Vamos aplicando definiciones

$A = \partial(\lambda f)(x) \quad \lambda \cdot \partial f(x) = B$

- $\xi \in A \rightarrow \forall z \quad \lambda f(z) \geq \lambda f(x) + \xi(z - x)$

$$f(z) \geq f(x) + \frac{\xi}{\lambda}(z - x)$$ Función encima de hiperplano. Definición de subgradiente.

$$\rightarrow \mu = \frac{\xi}{\lambda} \in \partial f(x) \rightarrow \xi = \lambda\mu \in \lambda \partial f(x)$$

- $\xi \in B \Rightarrow \xi = \lambda \cdot \mu \quad \mu \in \partial f(x)$

$$\Rightarrow \lambda \cdot f(z) \geq \lambda \cdot f(x) + \lambda\mu(z - x)$$

Dejo el hiperplano por debajo de la función λ.

$$\rightarrow \lambda \cdot \mu = \xi \in \partial(\lambda f)(x) = A.$$

Luis: 50%, Adrian: 50%

Exercise 18. *Prove that the $\epsilon-insensitive$ loss function $l_\epsilon(z) = max0, —z— \ \epsilon$ is convex. Give also its subgradient $\delta l_\epsilon(x)$ at any $x \in R$*

There are two properties we need to keep in mind

1. Any semispace $x : px(< | >)\alpha$ is a convex set.

2. The interception between convex sets is still convex.

Therefore any finite interception of semispaces is a convex set. l_ϵ can be seen as an interception of semispaces or as the solution to the following equation system:

$$\begin{cases} y + x \geq \epsilon \\ y + x \geq -\epsilon \\ y \geq 0 \end{cases} \tag{1}$$

So epi$(l_\epsilon) = (x, y) \in R^2 : (x, y) solves(P)$ and this means its convex.

Then we have to calculate all the subgradients for $x = \epsilon$, as we already know that $l_\epsilon(\epsilon) = 0$, we require each point ξ for which $l_\epsilon(z) \geq \xi, \forall z$. The constrains would be:

$$\begin{cases} z - \epsilon \geq \xi(z - \epsilon), z - \epsilon \iff \xi \leq 1 \\ 0 \geq \xi(z - \epsilon), z \in [-\epsilon, \epsilon] \iff \xi \leq 0 \end{cases} \tag{2}$$

We can infer by this that $\xi in[0, 1]$ and then all we have left is to calculate the subgradients when $x = -\epsilon$ and we come up with the following system of equations:

$$\begin{cases} -1, x < -\epsilon \\ [-1, 0], x = -\epsilon \\ 0, -\epsilon < x < \epsilon \\ [0, 1], x = \epsilon \\ 1, x > \epsilon \end{cases} \tag{3}$$

19

19) Calcula la proximal de

$$h(x) = \max[-x, 0]$$

Epígrafo convexo → f convexa

$$\lambda \cdot \partial h(x) = \begin{cases} -\lambda & x<0 \\ [-\lambda, 0] & x=0 \\ 0 & x>0 \end{cases}$$

$(I+\lambda\partial f)(x)$ → Para invertirla

2º

Proximal de h ⇒ $(I+\lambda\partial h)^{-1}(x) = \begin{cases} x+\lambda & \text{si } x<-\lambda \\ 0 & \text{si } -\lambda \le x \le 0 \\ x & \text{si } x>0 \end{cases}$

$$g(x) = \max[0, |x| - \varepsilon]$$

Epígrafo de g(x) convexo
↓
g(x) convexa

$$\partial g(x) = \begin{cases} -1 & \text{si } x<-\varepsilon \\ [-1, 0] & \text{si } x=-\varepsilon \\ 0 & \text{si } -\varepsilon<x<\varepsilon \\ [0, 1] & \text{si } x=\varepsilon \\ 1 & \text{si } x>\varepsilon \end{cases}$$

$$(I+\lambda\partial g(x)) = \begin{cases} -\lambda+x & \text{si } x<-\varepsilon \\ [-\lambda-\varepsilon, -\varepsilon] & \text{si } x=-\varepsilon \\ x & \text{si } -\varepsilon<x<\varepsilon \\ [\varepsilon, \lambda+\varepsilon] & \text{si } x=\varepsilon \\ \lambda+x & \text{si } x>\varepsilon \end{cases}$$

$\xrightarrow[\text{continua}]{\text{Graficamos}}$ *

Ejercicio 2E

Continúe

1° giro 90° (a la izquierda)

2° flip sobre eje y

Tras aplicar la transformación 1° y 2°, nos queda (para obtener el inverso)

$(I+\lambda\partial g)^{-1}(x)$

$$\text{proximal de } g(x) = (I+\lambda\partial g)^{-1}(x) = \begin{cases} x+\lambda & \text{si } x < -(\lambda+\varepsilon) \\ -\varepsilon & \text{si } -(\lambda+\varepsilon) \le x \le -\varepsilon \\ x & \text{si } -\varepsilon < x < \varepsilon \\ \varepsilon & \text{si } \varepsilon \le x \le (\lambda+\varepsilon) \\ x-\lambda & \text{si } x > (\lambda+\varepsilon) \end{cases}$$

Solución