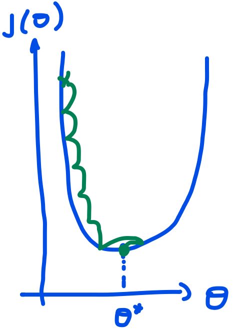
Análisis Matemático para Inteligencia Artificial

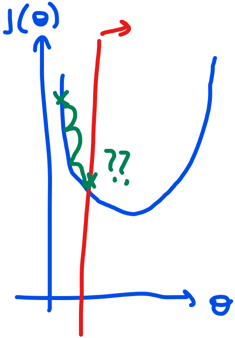
Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Optimización con restricciones

¿Qué pasa cuando el mínimo "clásico" no es un valor válido?





$\exists J(\theta) \ \forall \theta$ $\longrightarrow$ ya no se cumple

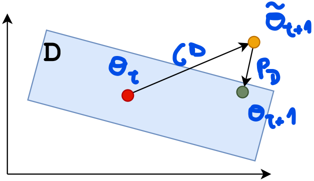
- mejorar $J(\theta)$
- no caer fuera de la región válida

Un "parche": Projected Gradient Descent

¿Cuál es el peligro de usar GD *as-is*? Caer afuera de la región válida D.

¿Cómo lo podemos corregir *"fácil"*? Buscamos el valor válido θ_{t+1} más cercano al update propuesto $\tilde{\theta}_{t+1} \rightarrow$ ¡GD + proyección ortogonal!

$$\theta_{t+1} = \Pi_D(\tilde{\theta}_{t+1}) = P_D \cdot (\theta_t - \gamma \cdot g_t)$$



Esto solo tiene sentido si proyectar es *barato*, pero a veces lo es.
Ejemplo: proyectar a valores no negativos es aplicar $\theta_{t+1} = max(\tilde{\theta}_{t+1}, 0)$.

Optimización con restricciones de igualdad

Definimos un problema de optimización con restricciones de igualdad en *formato estándar*:

$$\|\vec{x}\| = 1 \iff \|\vec{x}\|^2 = 1^2 = 1 \iff x_1^2 + x_2^2 + \ldots + x_n^2 = 1$$

$$\underbrace{x_1^2 + x_2^2 + \ldots + x_n^2 - 1}_{g_1(x_1, \ldots, x_n)} = 0$$

$$\min \; f(\overbrace{x_1, \ldots, x_n}^{n \text{ vars}})$$

$$\left.\begin{array}{r} \underset{\text{subject to}}{\text{s.t.}} \; g_1(x_1, \ldots, x_n) = 0 \\ \vdots \\ g_m(x_1, \ldots, x_n) = 0 \end{array}\right\} m \text{ restr.}$$

donde $f, g_i : \mathbb{R}^n \to \mathbb{R}$ con $i = 1, \ldots m$ están definidas sobre $R \subset \mathbb{R}^n$.

Se define la región válida $D = \{\vec{x} \in R : g_i(\vec{x}) = 0 \; \forall \, i = 1, \ldots, m\}$.

Se define el *Lagrangiano* del problema $\mathscr{L} : \mathbb{R}^{n+m} \to \mathbb{R}$:

$$\mathscr{L}(\underbrace{\vec{\lambda}}_{\text{multip. de Lagrange}}, \vec{x}) = f(\vec{x}) + \lambda_1 g_1(\vec{x}) + \ldots + \lambda_m g_m(\vec{x}) = f(\vec{x}) + \vec{\lambda} \cdot \vec{g}(\vec{x})$$

Optimización con restricciones de desigualdad

Si agregamos condiciones de desigualdad queda:

$$\begin{aligned} &\min\ f(\overbrace{x_1, ..., x_n}^{n \text{ vars.}}) \\ &s.t.\ g_i(x_1, ..., x_n) = 0 \ \Big\}\ l \\ &\quad h_j(x_1, ..., x_n) \leq 0 \ \Big\}\ k \end{aligned} \ \Bigg] \ m \text{ restr.}$$

con $i = 1, \ldots, l$ y $j = 1, \ldots, k$ suponiendo $l + k = m$.

Ahora tenemos que la región válida es

$$D = \{\vec{x} \in R : g_i(\vec{x}) = 0 \ \forall\, i = 1, \ldots, l \ \wedge h_j(\vec{x}) \leq 0 \ \forall\, j = 1, \ldots, k\}$$

Y el Lagrangiano es:

$$\mathscr{L}(\underbrace{\vec{\lambda}}_{l}, \underbrace{\vec{x}}_{n}, \underbrace{\vec{\mu}}_{k}) = f(\vec{x}) + \sum_{i=1}^{l} \lambda_i g_i(\vec{x}) + \sum_{j=1}^{k} \mu_j h_j(\vec{x}) = f(\vec{x}) + \vec{\lambda} \cdot \vec{\mathrm{g}}(\vec{x}) + \vec{\mu} \cdot \vec{\mathrm{h}}(\vec{x})$$

$= n+m$

σ^*

Sea un punto $\vec{x}^* \in D$ tal que $f, g_i, h_j \in \mathcal{C}^1(\mathcal{E}(\vec{x}^*))$. Bajo ciertas condiciones de regularidad, **si** $\vec{x}^*$ es un mínimo local **entonces** existen $\vec{\lambda}^* \in \mathbb{R}^l, \vec{\mu}^* \in \mathbb{R}^k$ tales que:

1. (Estacionariedad) $\nabla_{\vec{x}} \mathscr{L}(\vec{\lambda}^*, \vec{x}^*, \vec{\mu}^*) = \vec{0}$) ópt. local
2. (Factibilidad primal)) validez
   1. $g_i(\vec{x}^*) = 0 \quad \forall i$
   2. $h_j(\vec{x}^*) \leq 0 \quad \forall j$
3. (Factibilidad dual) $\mu_j^* \geq 0 \quad \forall j$) holgura
4. (Holgura complementaria) $\underbrace{\mu_j^* \cdot h_j(\vec{x}^*)}_{} = 0 \quad \forall j$

$\forall j \;\; \mu_j^* \geq 0 \wedge h_j(\vec{x}^*) \leq 0$

pero alguno de los debe ser 0

Ejemplo analítico

$$
\begin{aligned}
\min_{x,y} \quad & x^2 + y^2 \quad) \; f(x,y) \\
\text{s.t.} \quad & x + y = 1 \quad) \; g_1 \\
& x \geq 0 \quad) \; h_1 \Rightarrow h_1(x,y) = -x \leq 0
\end{aligned}
$$

El Lagrangiano es: $\mathscr{L}(x, y, \lambda, \mu) = x^2 + y^2 + \lambda(x + y - 1) + \mu(-x)$

Condiciones KKT:

	$\frac{\partial \mathcal{L}}{\partial x}$	$\frac{\partial \mathcal{L}}{\partial y}$
Estacionariedad:	$2x + \lambda - \mu = 0,$	$2y + \lambda = 0$
Factibilidad primal:	$x + y = 1,$	$x \geq 0$
Factibilidad dual:	$\mu \geq 0$	
Holgura complementaria:	$\mu x = 0$	

Solución:

$$x^* = 0.5,\ y^* = 0.5,\ \underbrace{\lambda^* = -1,\ \mu^* = 0}\ \Rightarrow f^* = 0.5$$