Análisis Matemático para Inteligencia Artificial

Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Optimización con restricciones

Motivación

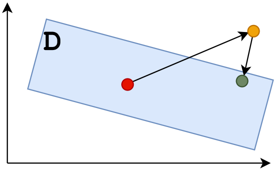
¿Qué pasa cuando el mínimo "clásico" no es un valor válido?

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*:

$$\begin{gathered}\min \; f(x_1, ..., x_n) \\ s.t. \; g_1(x_1, ..., x_n) = 0 \\ \vdots \\ g_m(x_1, ..., x_n) = 0\end{gathered}$$

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 $\mathscr{L} : \mathbb{R}^{n+m} \to \mathbb{R}$:

$$\mathscr{L}(\vec{\lambda}, \vec{x}) = f(\vec{x}) + \lambda_1 g_1(\vec{x}) + ... + \lambda_m g_m(\vec{x}) = f(\vec{x}) + \vec{\lambda} \cdot \vec{\mathrm{g}}(\vec{x})$$

Optimización con restricciones de desigualdad

Si agregamos condiciones de desigualdad queda:

$$\min \ f(x_1, ..., x_n)$$
$$s.t.\ g_i(x_1, ..., x_n) = 0$$
$$h_j(x_1, ..., x_n) \leq 0$$

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:

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

Condiciones necesarias de Karush-Kuhn-Tucker (KKT)

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

Ejemplo analítico

$$\min_{x,y} \quad x^2 + y^2$$
$$\text{s.t.} \quad x + y = 1$$
$$x \geq 0$$

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

Condiciones KKT:

Estacionariedad:	$2x + \lambda - \mu = 0, \quad 2y + \lambda = 0$
Factibilidad primal:	$x + y = 1, \quad x \geq 0$
Factibilidad dual:	$\mu \geq 0$
Holgura complementaria:	$\mu x = 0$

Solución:

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