Optimización y Probabilidad

Fernando Lozano

Universidad de los Andes

8 de agosto de 2014

Por qué optimización

Por qué optimización

- En el fondo de cualquier problema en ML siempre reside un problema de optimización

Por qué optimización

- En el fondo de cualquier problema en ML siempre reside un problema de optimización
- Escoger el mejor modelo (NN,SVM,...) de acuerdo a la meta de aprendizaje.

Por qué optimización

- En el fondo de cualquier problema en ML siempre reside un problema de optimización
- Escoger el mejor modelo (NN,SVM,. . .) de acuerdo a la meta de aprendizaje.
- Dos tipos:

Por qué optimización

- En el fondo de cualquier problema en ML siempre reside un problema de optimización
- Escoger el mejor modelo (NN,SVM,...) de acuerdo a la meta de aprendizaje.
- Dos tipos:
  1. Sin restricciones.

Por qué optimización

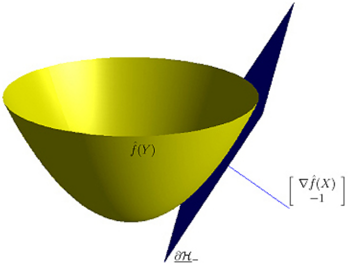
- En el fondo de cualquier problema en ML siempre reside un problema de optimización
- Escoger el mejor modelo (NN,SVM,...) de acuerdo a la meta de aprendizaje.
- Dos tipos:
  1. Sin restricciones.
  2. Con restricciones.

Por qué optimización

- En el fondo de cualquier problema en ML siempre reside un problema de optimización
- Escoger el mejor modelo (NN,SVM,...) de acuerdo a la meta de aprendizaje.
- Dos tipos:
  1. Sin restricciones.
  2. Con restricciones.
- Deseable:

Por qué optimización

- En el fondo de cualquier problema en ML siempre reside un problema de optimización
- Escoger el mejor modelo (NN,SVM,...) de acuerdo a la meta de aprendizaje.
- Dos tipos:
  1. Sin restricciones.
  2. Con restricciones.
- Deseable: convexidad.



Ejemplo: Regresión/Clasificación

- Suponemos que existe una relación funcional entre y y $\mathbf{x} = (x_1, x_2, \ldots, x_n)$

Ejemplo: Regresión/Clasificación

- Suponemos que existe una relación funcional entre y y $\mathbf{x} = (x_1, x_2, \ldots, x_n)$
- Conocemos mediciones $\{\mathbf{x}_i, y_i\}_{i=1}^{m}$ posiblemente ruidosas.

Ejemplo: Regresión/Clasificación

- Suponemos que existe una relación funcional entre y y $\mathbf{x} = (x_1, x_2, \ldots, x_n)$
- Conocemos mediciones $\{\mathbf{x}_i, y_i\}_{i=1}^{m}$ posiblemente ruidosas.
- Modelo paramétrico $y = f(\mathbf{x}, \boldsymbol{\Theta})$ ($\boldsymbol{\Theta}$: Vector de parámetros).

Ejemplo: Regresión/Clasificación

- Suponemos que existe una relación funcional entre y y $\mathbf{x} = (x_1, x_2, \ldots, x_n)$
- Conocemos mediciones $\{\mathbf{x}_i, y_i\}_{i=1}^m$ posiblemente ruidosas.
- Modelo paramétrico $y = f(\mathbf{x}, \boldsymbol{\Theta})$ ($\boldsymbol{\Theta}$: Vector de parámetros).
- Queremos resolver

$$\min_{\boldsymbol{\Theta}} \sum_{i=1}^{m} l(\mathbf{x}_i, y_i, f(\mathbf{x}_i, \boldsymbol{\Theta}))$$

Donde $l(.,.)$ es una función de error.

Ejemplo: Regresión/Clasificación

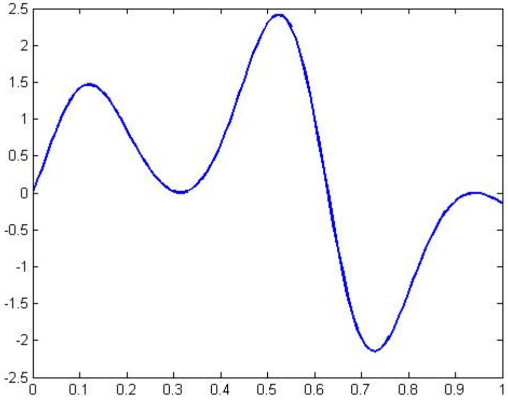
- Suponemos que existe una relación funcional entre y y $\mathbf{x} = (x_1, x_2, \ldots, x_n)$
- Conocemos mediciones $\{\mathbf{x}_i, y_i\}_{i=1}^m$ posiblemente ruidosas.
- Modelo paramétrico $y = f(\mathbf{x}, \boldsymbol{\Theta})$ ($\boldsymbol{\Theta}$: Vector de parámetros).
- Queremos resolver

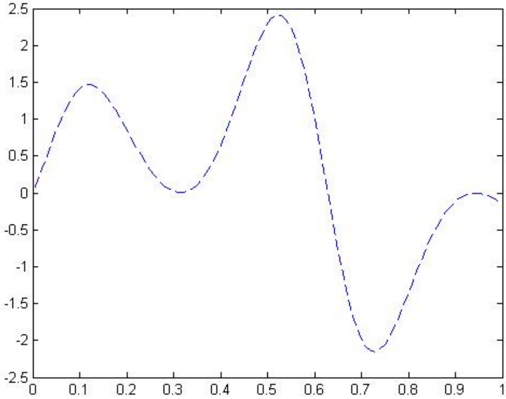
$$\min_{\boldsymbol{\Theta}} \sum_{i=1}^{m} l(\mathbf{x}_i, y_i, f(\mathbf{x}_i, \boldsymbol{\Theta}))$$

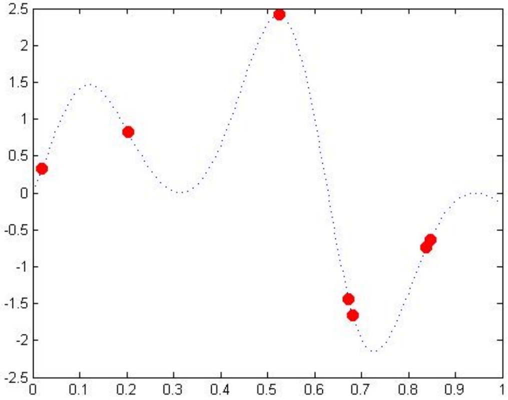
Donde $l(.,.)$ es una función de error.

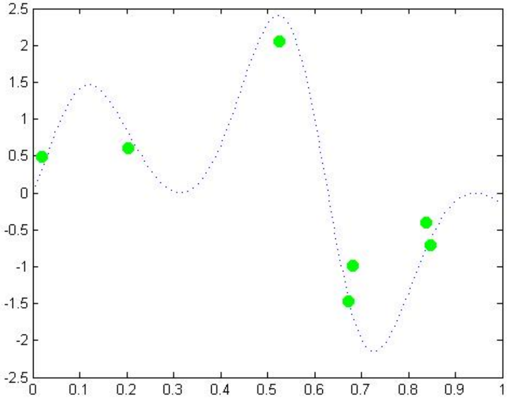
- Por ejemplo:

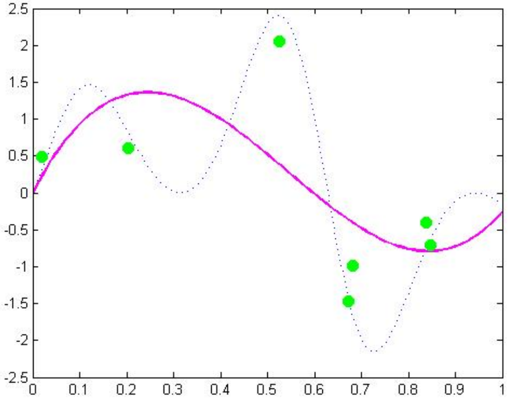
$$\min_{\boldsymbol{\Theta}} \sum_{i=1}^{m} (f(\mathbf{x_i}, \boldsymbol{\Theta}) - y_i)^2$$


2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Ejemplo: Máxima Verosimilitud

- Deseamos modelar distribución de probabilidad conjunta de vector aleatorio $\mathbf{x}$, usando observaciones $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.

Ejemplo: Máxima Verosimilitud

- Deseamos modelar distribución de probabilidad conjunta de vector aleatorio $\mathbf{x}$, usando observaciones $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.
- Asumiendo independencia, suponemos modelo de la forma

$$L(\boldsymbol{\Theta}) = \prod_{i=1}^{n} p(\mathbf{x}_i | \boldsymbol{\Theta})$$

Ejemplo: Máxima Verosimilitud

- Deseamos modelar distribución de probabilidad conjunta de vector aleatorio $\mathbf{x}$, usando observaciones $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.
- Asumiendo independencia, suponemos modelo de la forma

$$L(\boldsymbol{\Theta}) = \prod_{i=1}^{n} p(\mathbf{x}_i | \boldsymbol{\Theta})$$

Queremos resolver:

$$\max_{\boldsymbol{\Theta}} \{\log L(\boldsymbol{\Theta})\}$$

Por ejemplo:

- $\mathbf{x}$ con distribución Normal, parámetro es media $\boldsymbol{\mu}$,

$$p(\mathbf{x}|\boldsymbol{\mu}) = \frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

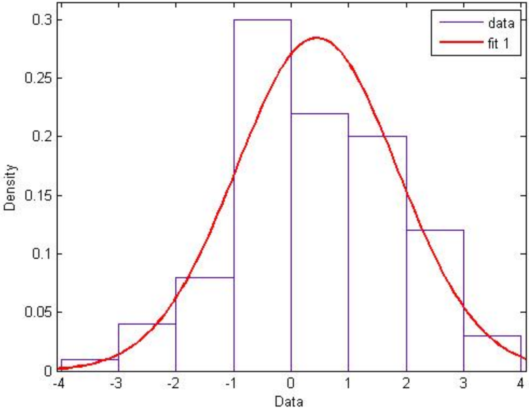
Por ejemplo:

- $\mathbf{x}$ con distribución Normal, parámetro es media $\boldsymbol{\mu}$,

$$p(\mathbf{x}|\boldsymbol{\mu}) = \frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

- Queremos resolver:

$$\min_{\boldsymbol{\mu}} \sum_{i=1}^{n} \left((\mathbf{x}_i-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})\right)$$


data
fit 1
Density
0.3
0.25
0.2
0.15
0.1
0.05
0
-4
-3
-2
-1
0
1
2
3
4
Data

Optimización sin restricciones

- Problema general:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \quad f(\mathbf{x})$$

Optimización sin restricciones

- Problema general:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \quad f(\mathbf{x})$$

- Deseable que f sea una función suave: derivadas de primer y segundo orden existen y son continuas.

Optimización sin restricciones

- Problema general:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \quad f(\mathbf{x})$$

- Deseable que f sea una función suave: derivadas de primer y segundo orden existen y son continuas.
- Mínimo local vs. Mínimo global.

Optimización sin restricciones

- Problema general:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \quad f(\mathbf{x})$$

- Deseable que f sea una función suave: derivadas de primer y segundo orden existen y son continuas.
- Mínimo local vs. Mínimo global.
- Convexidad.

Derivadas de primer y segundo orden

Gradiente:

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

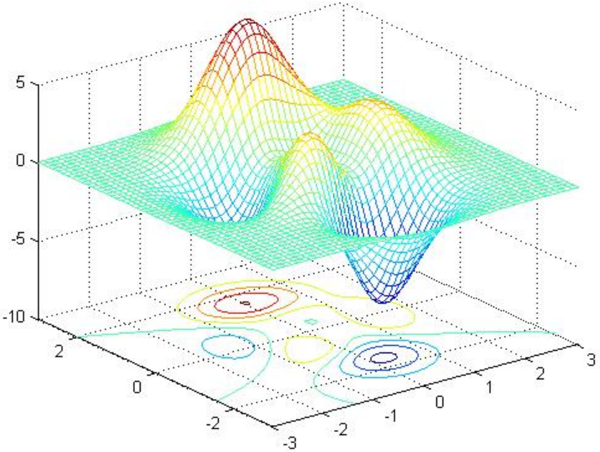
Derivadas de primer y segundo orden

Gradiente:

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

Hessiana:

$$\nabla^2_{\mathbf{x}} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{bmatrix}$$


5
0
-5
-10
2
0
-2
-3
-2
-1
0
1
2
3

Algoritmo iterativo de descenso

Incialice $\mathbf{x}_0$

Algoritmo iterativo de descenso

Inicialice $\mathbf{x}_0$
repeat

Algoritmo iterativo de descenso

Inicialice $\mathbf{x}_0$
repeat
  Escoja dirección $\mathbf{d}$

Algoritmo iterativo de descenso

Inicialice $\mathbf{x}_0$
repeat
  Escoja dirección $\mathbf{d}$
  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}$

Algoritmo iterativo de descenso

Inicialice $\mathbf{x}_0$
repeat
  Escoja dirección $\mathbf{d}$
  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}$
until Condición de terminación.

Algoritmo iterativo de descenso

Inicialice $\mathbf{x}_0$
repeat
  Escoja dirección $\mathbf{d}$
  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}$
until Condición de terminación.

Escogencia de α_k

Escogencia de α_k

- Constante

Escogencia de α_k

- Constante (tasa de aprendizaje)

Escogencia de α_k

- Constante (tasa de aprendizaje)
- Idealmente $\alpha_k = \arg\min_\alpha f\left(\mathbf{x}_k + \alpha\mathbf{d}\right)$

Escogencia de α_k

- Constante (tasa de aprendizaje)
- Idealmente $\alpha_k = \arg\min_\alpha f\left(\mathbf{x}_k + \alpha\mathbf{d}\right) \Rightarrow$ Algoritmos de búsqueda de línea.

Escogencia de α_k

- Constante (tasa de aprendizaje)
- Idealmente $\alpha_k = \arg\min_\alpha f(\mathbf{x}_k + \alpha \mathbf{d}) \Rightarrow$ Algoritmos de búsqueda de línea.
  1. Golden Search, Fibonacci.
  2. Newton, ajuste cúbico.
  3. Backtracking.
  4. $\vdots$

Escogencia de **d**

Escogencia de d

- Descenso de gradiente:

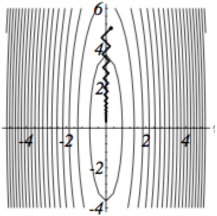
$$\mathbf{d}_k = -\nabla_{\mathbf{x}} f(\mathbf{x}_k)$$

Escogencia de **d**

- Descenso de gradiente:

$$\mathbf{d}_k = -\nabla_{\mathbf{x}} f(\mathbf{x}_k)$$

  - Convergencia lenta.

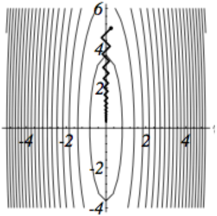


Escogencia de **d**

- Descenso de gradiente:

$$\mathbf{d}_k = -\nabla_{\mathbf{x}} f(\mathbf{x}_k)$$

  - Convergencia lenta.



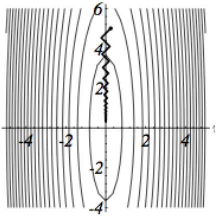
  - Simple: sólo primeras derivadas.

Escogencia de **d**

- Descenso de gradiente:

$$\mathbf{d}_k = -\nabla_{\mathbf{x}} f(\mathbf{x}_k)$$

  - Convergencia lenta.



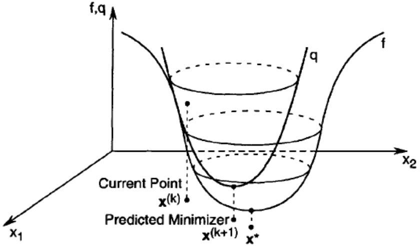
  - Simple: sólo primeras derivadas.
  - Version estocástica/ en línea.

- Método de Newton:

$$\mathbf{d}_k = -(\nabla^2_{\mathbf{x}} f(\mathbf{x}_k))^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$$

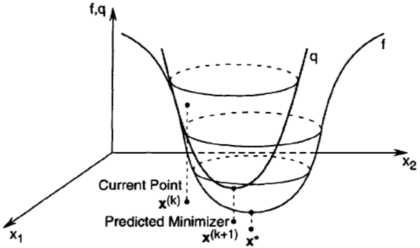
- Método de Newton:

$$\mathbf{d}_k = -(\nabla^2_{\mathbf{x}} f(\mathbf{x}_k))^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$$



- Método de Newton:

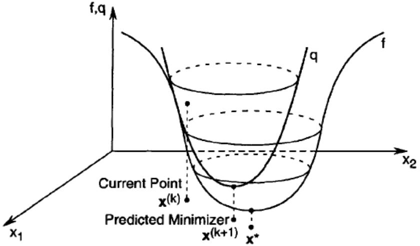
$$\mathbf{d}_k = -(\nabla^2_{\mathbf{x}} f(\mathbf{x}_k))^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$$



  - Convergencia rápida (orden 2).

- Método de Newton:

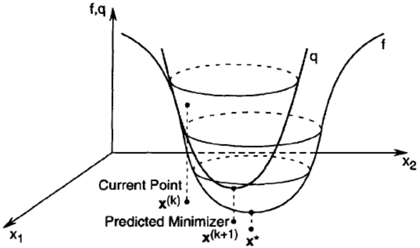
$$\mathbf{d}_k = -(\nabla^2_{\mathbf{x}} f(\mathbf{x}_k))^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$$



  - Convergencia rápida (orden 2).
  - Requiere cálculo de $(\nabla^2_{\mathbf{x}} f(\mathbf{x}_k))^{-1}$

- Método de Newton:

$$\mathbf{d}_k = -(\nabla^2_{\mathbf{x}} f(\mathbf{x}_k))^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}_k)$$



  - Convergencia rápida (orden 2).
  - Requiere cálculo de $(\nabla^2_{\mathbf{x}} f(\mathbf{x}_k))^{-1}$
  - Frágil en el caso no convexo.

- Métodos de Quasi-Newton (BFGS,DFP).

- Métodos de Quasi-Newton (BFGS,DFP).
    - Aproximación sucesiva de $(\nabla^2_{\mathbf{x}} f(\mathbf{x}_k))^{-1}$ en términos de primeras derivadas.

- Métodos de Quasi-Newton (BFGS,DFP).
  - Aproximación sucesiva de $(\nabla^2_{\mathbf{x}} f(\mathbf{x}_k))^{-1}$ en términos de primeras derivadas.
  - Método preferido en problemas en mediana/gran escala.

- Métodos de Quasi-Newton (BFGS,DFP).
  - Aproximación sucesiva de $(\nabla^2_{\mathbf{x}} f(\mathbf{x}_k))^{-1}$ en términos de primeras derivadas.
  - Método preferido en problemas en mediana/gran escala.
- Gradiente conjugado.

- Métodos de Quasi-Newton (BFGS,DFP).
  - Aproximación sucesiva de $(\nabla^2_{\mathbf{x}} f(\mathbf{x}_k))^{-1}$ en términos de primeras derivadas.
  - Método preferido en problemas en mediana/gran escala.
- Gradiente conjugado.
  - Versión mejorada de descenso de gradiente a bajo costo computacional.

Problemas de Optimizacion con Restricciones

$$\min_{\mathbf{x}\in\Omega} \quad f(\mathbf{x})$$

$$\text{sujeto a}$$

$$g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, p$$

$$h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m$$

Problemas de Optimizacion con Restricciones

$$\min_{\mathbf{x}\in\Omega} \quad f(\mathbf{x})$$

$$\text{sujeto a}$$

$$g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, p$$

$$h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m$$

- $g_i(\mathbf{x})$ Restricciones de desigualdad.
- $h_i(\mathbf{x})$ Restricciones de igualdad.

Problemas de Optimizacion con Restricciones

$$\min_{\mathbf{x}\in\Omega} \quad f(\mathbf{x})$$

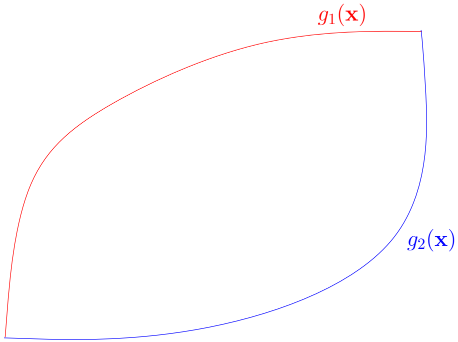
$$\text{sujeto a}$$

$$g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, p$$

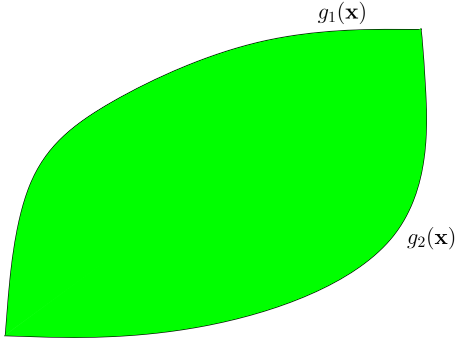
$$h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m$$

- $g_i(\mathbf{x})$ Restricciones de desigualdad.
- $h_i(\mathbf{x})$ Restricciones de igualdad.
- $\mathbf{x}^*$: Solución óptima.

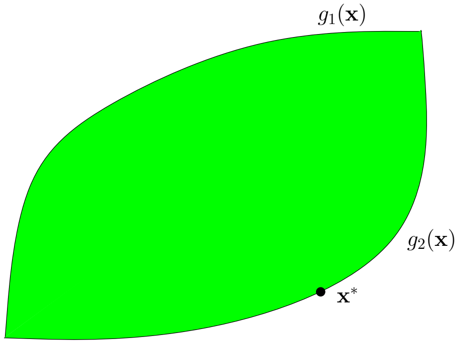
Restricciones Activas



Restricciones Activas

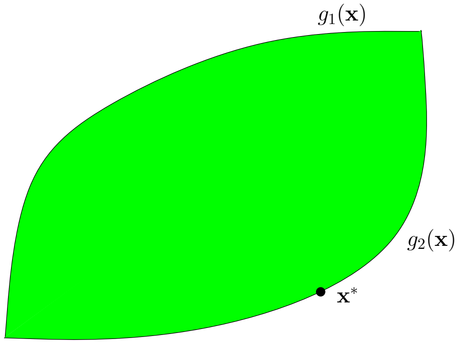


Restricciones Activas



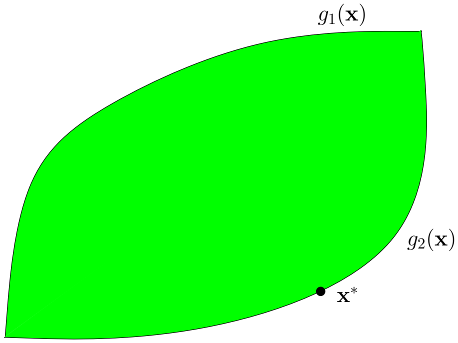
- $g_2(\mathbf{x})$ es activa en $\mathbf{x}^*$.

Restricciones Activas



- $g_2(\mathbf{x})$ es activa en $\mathbf{x}^*$.
- Restricciones inactivas no influyen en condiciones para un mínimo local.

Restricciones Activas



- $g_2(\mathbf{x})$ es activa en $\mathbf{x}^*$.
- Restricciones inactivas no influyen en condiciones para un mínimo local.
- Restricciones de igualdad son siempre activas en un punto factible.

El Langrangiano

$$\min_{\mathbf{x}\in\Omega} \quad f_0(\mathbf{x})$$

$$\text{sujeto a}$$

$$f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m$$

$$h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p$$

El Langrangiano

$$\begin{aligned} \min_{\mathbf{x}\in\Omega} \quad & f_0(\mathbf{x}) \\ \text{sujeto a} & \\ f_i(\mathbf{x}) \leq 0, \quad & i = 1, \ldots, m \\ h_i(\mathbf{x}) = 0, \quad & i = 1, \ldots, p \end{aligned}$$

- El Lagrangiano $L : (\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p) \to \mathbb{R}$,

El Langrangiano

$$\min_{\mathbf{x}\in\Omega} \quad f_0(\mathbf{x})$$

sujeto a

$$f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m$$

$$h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p$$

- El Lagrangiano $L : (\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p) \to \mathbb{R}$,

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x})$$

El Langrangiano

$$\min_{\mathbf{x}\in\Omega} \quad f_0(\mathbf{x})$$

$$\text{sujeto a}$$

$$f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m$$

$$h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p$$

- El Lagrangiano $L : (\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p) \to \mathbb{R}$,

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x})$$

El Langrangiano

$$\begin{aligned} \min_{\mathbf{x}\in\Omega} \quad & f_0(\mathbf{x}) \\ \text{sujeto a} & \\ f_i(\mathbf{x}) \leq 0, \quad & i = 1, \dots, m \\ h_i(\mathbf{x}) = 0, \quad & i = 1, \dots, p \end{aligned}$$

- El Lagrangiano $L : (\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p) \to \mathbb{R}$,

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$$

El Langrangiano

$$\min_{\mathbf{x}\in\Omega} \quad f_0(\mathbf{x})$$

sujeto a

$$f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m$$

$$h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p$$

- El Lagrangiano $L : (\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p) \to \mathbb{R}$,

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$$

- $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_m]$ y $\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_p]$ son las variables duales o multiplicadores de Lagrange.

La función Dual

- Definimos la función dual como el valor mínimo del Lagrangiano sobre $\mathbf{x}$:

La función Dual

- Definimos la función dual como el valor mínimo del Lagrangiano sobre $\mathbf{x}$:

$$\begin{aligned} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\ &= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \right) \end{aligned}$$

La función Dual

- Definimos la función dual como el valor mínimo del Lagrangiano sobre $\mathbf{x}$:

$$
\begin{aligned}
g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\
&= \min_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \right)
\end{aligned}
$$

La función Dual

- Definimos la función dual como el valor mínimo del Lagrangiano sobre $\mathbf{x}$:

$$
\begin{aligned}
g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\
&= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \right)
\end{aligned}
$$

La función Dual

- Definimos la función dual como el valor mínimo del Lagrangiano sobre $\mathbf{x}$:

$$\begin{aligned} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\ &= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \right) \end{aligned}$$

- $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ es una función cóncava.

La función dual es una cota inferior de $f(\mathbf{x}^*)$

- Sea $f(\mathbf{x}^*) = p^*$.

La función dual es una cota inferior de $f(\mathbf{x}^*)$

- Sea $f(\mathbf{x}^*) = p^*$. Para cualquier $\boldsymbol{\lambda} \geq \mathbf{0}$ y cualquier $\boldsymbol{\nu}$ tenemos:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$$

La función dual es una cota inferior de $f(\mathbf{x}^*)$

- Sea $f(\mathbf{x}^*) = p^*$. Para cualquier $\boldsymbol{\lambda} \geq \mathbf{0}$ y cualquier $\boldsymbol{\nu}$ tenemos:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$$

- Suponga que $\mathbf{x}^\circ$ es un punto factible y $\boldsymbol{\lambda} \geq \mathbf{0}$, entonces

La función dual es una cota inferior de $f(\mathbf{x}^*)$

- Sea $f(\mathbf{x}^*) = p^*$. Para cualquier $\boldsymbol{\lambda} \geq \mathbf{0}$ y cualquier $\boldsymbol{\nu}$ tenemos:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$$

- Suponga que $\mathbf{x}^\circ$ es un punto factible y $\boldsymbol{\lambda} \geq \mathbf{0}$, entonces

$$\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}^\circ) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}^\circ) \leq 0$$

La función dual es una cota inferior de $f(\mathbf{x}^*)$

- Sea $f(\mathbf{x}^*) = p^*$. Para cualquier $\boldsymbol{\lambda} \geq \mathbf{0}$ y cualquier $\boldsymbol{\nu}$ tenemos:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$$

- Suponga que $\mathbf{x}^\circ$ es un punto factible y $\boldsymbol{\lambda} \geq \mathbf{0}$, entonces

$$\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}^\circ) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}^\circ) \leq 0$$

$$f_0(\mathbf{x}^\circ) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}^\circ) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}^\circ) \leq f_0(\mathbf{x}^\circ)$$

La función dual es una cota inferior de $f(\mathbf{x}^*)$

- Sea $f(\mathbf{x}^*) = p^*$. Para cualquier $\boldsymbol{\lambda} \geq \mathbf{0}$ y cualquier $\boldsymbol{\nu}$ tenemos:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$$

- Suponga que $\mathbf{x}^\circ$ es un punto factible y $\boldsymbol{\lambda} \geq \mathbf{0}$, entonces

$$\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}^\circ) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}^\circ) \leq 0$$

$$f_0(\mathbf{x}^\circ) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}^\circ) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}^\circ) \leq f_0(\mathbf{x}^\circ)$$

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq L(\mathbf{x}^\circ, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\mathbf{x}^\circ)$$

La función dual es una cota inferior de $f(\mathbf{x}^*)$

- Sea $f(\mathbf{x}^*) = p^*$. Para cualquier $\boldsymbol{\lambda} \geq \mathbf{0}$ y cualquier $\boldsymbol{\nu}$ tenemos:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$$

- Suponga que $\mathbf{x}^\circ$ es un punto factible y $\boldsymbol{\lambda} \geq \mathbf{0}$, entonces

$$\begin{aligned}
\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}^\circ) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}^\circ) &\leq 0 \\
f_0(\mathbf{x}^\circ) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}^\circ) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}^\circ) &\leq f_0(\mathbf{x}^\circ) \\
g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq L(\mathbf{x}^\circ, \boldsymbol{\lambda}, \boldsymbol{\nu}) &\leq f_0(\mathbf{x}^\circ)
\end{aligned}$$

- Podemos reemplazar $\mathbf{x}^\circ$ por $\mathbf{x}^*$.

La función dual es una cota inferior de $f(\mathbf{x}^*)$

- Sea $f(\mathbf{x}^*) = p^*$. Para cualquier $\boldsymbol{\lambda} \geq \mathbf{0}$ y cualquier $\boldsymbol{\nu}$ tenemos:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$$

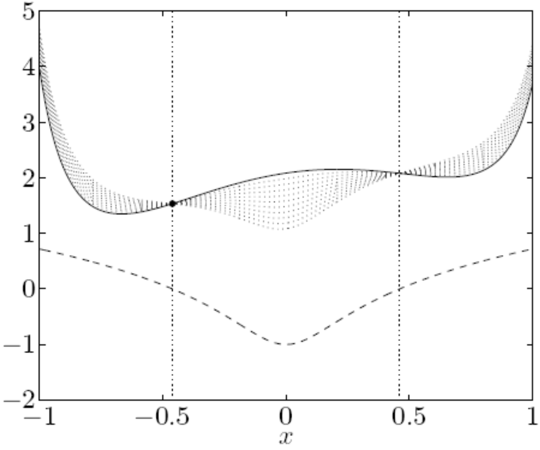
- Suponga que $\mathbf{x}^\circ$ es un punto factible y $\boldsymbol{\lambda} \geq \mathbf{0}$, entonces

$$\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}^\circ) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}^\circ) \leq 0$$

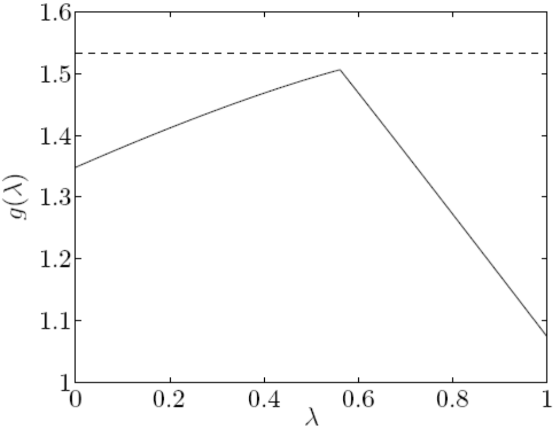
$$f_0(\mathbf{x}^\circ) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}^\circ) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}^\circ) \leq f_0(\mathbf{x}^\circ)$$

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq L(\mathbf{x}^\circ, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\mathbf{x}^\circ)$$

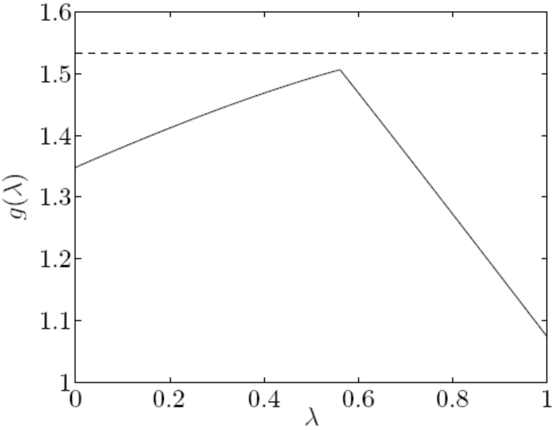
- Podemos reemplazar $\mathbf{x}^\circ$ por $\mathbf{x}^*$.
- Cota no trivial sólo si $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$



5
4
3
2
1
0
−1
−2
−1
−0.5
0
0.5
1
x



1.6
1.5
1.4
1.3
1.2
1.1
1
$g(\lambda)$
0
0.2
0.4
0.6
0.8
1
λ



- Note que aunque f_0 y f_i no son convexas, g es cóncava.

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

- El Lagrangiano:

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

- El Lagrangiano:

$$L(\mathbf{x}, \boldsymbol{\nu})$$

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

- El Lagrangiano:

$$L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{x}^T\mathbf{x} +$$

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

- El Lagrangiano:

$$L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{x}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

- El Lagrangiano:

$$L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{x}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

- Podemos hallar la función dual minimizando con respecto a $\mathbf{x}$:

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

- El Lagrangiano:

$$L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{x}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

- Podemos hallar la función dual minimizando con respecto a $\mathbf{x}$:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) =$$

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

- El Lagrangiano:

$$L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{x}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

- Podemos hallar la función dual minimizando con respecto a $\mathbf{x}$:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = 2\mathbf{x} + \mathbf{A}^T\boldsymbol{\nu}$$

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

- El Lagrangiano:

$$L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{x}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

- Podemos hallar la función dual minimizando con respecto a $\mathbf{x}$:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = 2\mathbf{x} + \mathbf{A}^T\boldsymbol{\nu} = 0$$

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

- El Lagrangiano:

$$L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{x}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

- Podemos hallar la función dual minimizando con respecto a $\mathbf{x}$:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = 2\mathbf{x} + \mathbf{A}^T\boldsymbol{\nu} = 0 \Rightarrow \mathbf{x} = -\frac{1}{2}\mathbf{A}^T\boldsymbol{\nu}$$

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

- El Lagrangiano:

$$L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{x}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

- Podemos hallar la función dual minimizando con respecto a $\mathbf{x}$:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = 2\mathbf{x} + \mathbf{A}^T\boldsymbol{\nu} = 0 \Rightarrow \mathbf{x} = -\frac{1}{2}\mathbf{A}^T\boldsymbol{\nu}$$

Reemplazando:

$$g(\boldsymbol{\nu}) = L(-\frac{1}{2}\mathbf{A}^T\boldsymbol{\nu}, \boldsymbol{\nu})$$

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

- El Lagrangiano:

$$L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{x}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

- Podemos hallar la función dual minimizando con respecto a $\mathbf{x}$:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = 2\mathbf{x} + \mathbf{A}^T\boldsymbol{\nu} = 0 \Rightarrow \mathbf{x} = -\frac{1}{2}\mathbf{A}^T\boldsymbol{\nu}$$

  Reemplazando:

$$g(\boldsymbol{\nu}) = L(-\frac{1}{2}\mathbf{A}^T\boldsymbol{\nu}, \boldsymbol{\nu}) = -\frac{1}{4}\boldsymbol{\nu}^T(\mathbf{A}\mathbf{A}^T)\boldsymbol{\nu} - \mathbf{b}^T\nu$$

Ejemplo

$$\min \mathbf{x}^T\mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p\times n}$$

- El Lagrangiano:

$$L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{x}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

- Podemos hallar la función dual minimizando con respecto a $\mathbf{x}$:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = 2\mathbf{x} + \mathbf{A}^T\boldsymbol{\nu} = 0 \Rightarrow \mathbf{x} = -\frac{1}{2}\mathbf{A}^T\boldsymbol{\nu}$$

  Reemplazando:

$$g(\boldsymbol{\nu}) = L(-\frac{1}{2}\mathbf{A}^T\boldsymbol{\nu}, \boldsymbol{\nu}) = -\frac{1}{4}\boldsymbol{\nu}^T(\mathbf{A}\mathbf{A}^T)\boldsymbol{\nu} - \mathbf{b}^T\nu$$

- Función cuadrática cóncava.

El Problema Dual de Lagrange

El Problema Dual de Lagrange

- Para cada $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ con $\boldsymbol{\lambda} > \mathbf{0}$, la función dual de Lagrange nos da una cota inferior de p^*.

El Problema Dual de Lagrange

- Para cada $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ con $\boldsymbol{\lambda} > \mathbf{0}$, la función dual de Lagrange nos da una cota inferior de p^*.
- Cuál es la mejor cota inferior que podemos obtener?

El Problema Dual de Lagrange

- Para cada $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ con $\boldsymbol{\lambda} > \mathbf{0}$, la función dual de Lagrange nos da una cota inferior de p^*.
- Cuál es la mejor cota inferior que podemos obtener?
- Problema de optimización:

$$\begin{aligned} \text{máx} \quad & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{sujeto a} \quad & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned}$$

El Problema Dual de Lagrange

- Para cada $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ con $\boldsymbol{\lambda} > \mathbf{0}$, la función dual de Lagrange nos da una cota inferior de p^*.
- Cuál es la mejor cota inferior que podemos obtener?
- Problema de optimización:

$$\begin{aligned} \text{máx} \quad & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{sujeto a} \quad & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned}$$

- Este es el problema dual de Lagrange asociado al problema primal original.

El Problema Dual de Lagrange

- Para cada $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ con $\boldsymbol{\lambda} > \mathbf{0}$, la función dual de Lagrange nos da una cota inferior de p^*.
- Cuál es la mejor cota inferior que podemos obtener?
- Problema de optimización:

$$\begin{array}{rl} \text{máx} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{sujeto a} & \boldsymbol{\lambda} \geq \mathbf{0} \end{array}$$

- Este es el problema dual de Lagrange asociado al problema primal original.
- $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ es factible en el dual si $\boldsymbol{\lambda} \geq \mathbf{0}$ y $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$.

El Problema Dual de Lagrange

- Para cada $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ con $\boldsymbol{\lambda} > \mathbf{0}$, la función dual de Lagrange nos da una cota inferior de p^*.
- Cuál es la mejor cota inferior que podemos obtener?
- Problema de optimización:

$$\begin{array}{rl} \text{máx} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{sujeto a} & \boldsymbol{\lambda} \geq \mathbf{0} \end{array}$$

- Este es el problema dual de Lagrange asociado al problema primal original.
- $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ es factible en el dual si $\boldsymbol{\lambda} \geq \mathbf{0}$ y $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$.
- SoluciÛn Ûptima dual $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ (multiplicadores de Lagrange óptimos).

Ejemplo

- Primal:

$$\min_{\mathbf{x}} \mathbf{x}^T \mathbf{x} \quad \text{sujeto a} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{p \times n}$$

- Dual

$$\max_{\boldsymbol{\nu}} -\frac{1}{4}\boldsymbol{\nu}(\mathbf{A}\mathbf{A}^T)\boldsymbol{\nu} - \mathbf{b}^T \nu$$

Dualidad Débil y Dualidad Fuerte

Dualidad Débil y Dualidad Fuerte

- Sea $d^* = g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, entonces,

Dualidad Débil y Dualidad Fuerte

- Sea $d^* = g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, entonces,

$$d^* \leq p^*$$

Dualidad Débil y Dualidad Fuerte

- Sea $d^* = g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, entonces,

$$d^* \leq p^*$$

- $p^* - d^*$ es la brecha de dualidad óptima.

Dualidad Débil y Dualidad Fuerte

- Sea $d^* = g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, entonces,

$$d^* \leq p^*$$

- $p^* - d^*$ es la brecha de dualidad óptima.
- Si la brecha de dualidad óptima es cero $(d^* = p^*)$ tenemos dualidad fuerte.

Dualidad Débil y Dualidad Fuerte

- Sea $d^* = g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, entonces,

$$d^* \leq p^*$$

- $p^* - d^*$ es la brecha de dualidad óptima.
- Si la brecha de dualidad óptima es cero $(d^* = p^*)$ tenemos dualidad fuerte.
- Dualidad fuerte: condiciones de Karush-Kuhn-Tucker.

Dualidad Débil y Dualidad Fuerte

- Sea $d^* = g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, entonces,

$$d^* \leq p^*$$

- $p^* - d^*$ es la brecha de dualidad óptima.
- Si la brecha de dualidad óptima es cero $(d^* = p^*)$ tenemos dualidad fuerte.
- Dualidad fuerte: condiciones de Karush-Kuhn-Tucker.
- Métodos de solución.

Dualidad Débil y Dualidad Fuerte

- Sea $d^* = g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, entonces,

$$d^* \leq p^*$$

- $p^* - d^*$ es la brecha de dualidad óptima.
- Si la brecha de dualidad óptima es cero $(d^* = p^*)$ tenemos dualidad fuerte.
- Dualidad fuerte: condiciones de Karush-Kuhn-Tucker.
- Métodos de solución.
- Tenemos dualidad fuerte, por ejemplo cuando se minimiza una función convexa en un poliedro convexo.

Holgura Complementaria

- Suponga que tenemos $p^* = d^*$.

Holgura Complementaria

- Suponga que tenemos $p^* = d^*$.
- Tenemos:

$$f_0(\mathbf{x}^*) = g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$$

Holgura Complementaria

- Suponga que tenemos $p^* = d^*$.
- Tenemos:

$$
\begin{aligned}
f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}) \right)
\end{aligned}
$$

Holgura Complementaria

- Suponga que tenemos $p^* = d^*$.
- Tenemos:

$$
\begin{aligned}
f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}) \right) \\
&\leq f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*)
\end{aligned}
$$

Holgura Complementaria

- Suponga que tenemos $p^* = d^*$.
- Tenemos:

$$
\begin{aligned}
f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}) \right) \\
&\leq f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) \\
&\leq f_0(\mathbf{x}^*)
\end{aligned}
$$

Holgura Complementaria

- Suponga que tenemos $p^* = d^*$.
- Tenemos:

$$\begin{aligned} f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\ &= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}) \right) \\ &\leq f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) \\ &\leq f_0(\mathbf{x}^*) \end{aligned}$$

Holgura Complementaria

- Suponga que tenemos $p^* = d^*$.
- Tenemos:

$$
\begin{aligned}
f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}) \right) \\
&= f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) \\
&= f_0(\mathbf{x}^*)
\end{aligned}
$$

Holgura Complementaria

- Suponga que tenemos $p^* = d^*$.
- Tenemos:

$$
\begin{aligned}
f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}) \right) \\
&= f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) \\
&= f_0(\mathbf{x}^*)
\end{aligned}
$$

- $\mathbf{x}^*$ minimiza $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ sobre $\mathbf{x}$.

Holgura Complementaria

- Suponga que tenemos $p^* = d^*$.
- Tenemos:

$$\begin{aligned} f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\ &= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}) \right) \\ &= f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) \\ &= f_0(\mathbf{x}^*) \end{aligned}$$

- $\mathbf{x}^*$ minimiza $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ sobre $\mathbf{x}$.
- $\sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) = 0$

- Podemos expresar la condición de holgura complementaria:

$$\sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) = 0$$

como

- Podemos expresar la condición de holgura complementaria:

$$\sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) = 0$$

como

$$\lambda_i^* > 0 \Rightarrow f_i(\mathbf{x}^*) = 0$$

- Podemos expresar la condición de holgura complementaria:

$$\sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) = 0$$

como

$$\lambda_i^* > 0 \Rightarrow f_i(\mathbf{x}^*) = 0$$

o

$$f_i(\mathbf{x}^*) < 0 \Rightarrow \lambda_i^* = 0$$

- Podemos expresar la condición de holgura complementaria:

$$\sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) = 0$$

como

$$\lambda_i^* > 0 \Rightarrow f_i(\mathbf{x}^*) = 0$$

o

$$f_i(\mathbf{x}^*) < 0 \Rightarrow \lambda_i^* = 0$$

- Es decir, el i-ésimo multiplicador de Lagrange es cero, a no ser que la restricción correspondiente sea activa

Condiciones de Karush-Khun-Tucker (KKT)

- Suponga $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ diferenciables.

Condiciones de Karush-Khun-Tucker (KKT)

- Suponga $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ diferenciables.
- En un par de puntos óptimos para el primal $(\mathbf{x}^*)$ y el dual $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, se cumplen las siguientes condiciones:

Condiciones de Karush-Khun-Tucker (KKT)

- Suponga $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ diferenciables.
- En un par de puntos óptimos para el primal $(\mathbf{x}^*)$ y el dual $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, se cumplen las siguientes condiciones:

$$f_i(\mathbf{x}^*) \leq 0, \quad i = 1, \ldots, m.$$

Condiciones de Karush-Khun-Tucker (KKT)

- Suponga $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ diferenciables.
- En un par de puntos óptimos para el primal $(\mathbf{x}^*)$ y el dual $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, se cumplen las siguientes condiciones:

$$\begin{aligned} f_i(\mathbf{x}^*) &\leq 0, \quad i = 1, \ldots, m. \\ h_i(\mathbf{x}^*) &= 0, \quad i = 1, \ldots, p. \end{aligned}$$

Condiciones de Karush-Khun-Tucker (KKT)

- Suponga $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ diferenciables.
- En un par de puntos óptimos para el primal $(\mathbf{x}^*)$ y el dual $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, se cumplen las siguientes condiciones:

$$\begin{aligned} f_i(\mathbf{x}^*) &\leq 0, \quad i = 1, \ldots, m. \\ h_i(\mathbf{x}^*) &= 0, \quad i = 1, \ldots, p. \\ \lambda_i^* &\geq 0, \quad i = 1, \ldots, m. \end{aligned}$$

Condiciones de Karush-Khun-Tucker (KKT)

- Suponga $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ diferenciables.
- En un par de puntos óptimos para el primal $(\mathbf{x}^*)$ y el dual $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, se cumplen las siguientes condiciones:

$$
\begin{aligned}
f_i(\mathbf{x}^*) &\leq 0, \quad i = 1, \ldots, m. \\
h_i(\mathbf{x}^*) &= 0, \quad i = 1, \ldots, p. \\
\lambda_i^* &\geq 0, \quad i = 1, \ldots, m. \\
\lambda_i^* f_i(\mathbf{x}^*) &= 0, \quad i = 1, \ldots, m.
\end{aligned}
$$

Condiciones de Karush-Khun-Tucker (KKT)

- Suponga $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ diferenciables.
- En un par de puntos óptimos para el primal $(\mathbf{x}^*)$ y el dual $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, se cumplen las siguientes condiciones:

$$
\begin{aligned}
f_i(\mathbf{x}^*) &\leq 0, \quad i = 1, \ldots, m. \\
h_i(\mathbf{x}^*) &= 0, \quad i = 1, \ldots, p. \\
\lambda_i^* &\geq 0, \quad i = 1, \ldots, m. \\
\lambda_i^* f_i(\mathbf{x}^*) &= 0, \quad i = 1, \ldots, m.
\end{aligned}
$$

$$\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(\mathbf{x}^*) = 0$$

Condiciones de Karush-Khun-Tucker (KKT)

- Suponga $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ diferenciables.
- En un par de puntos óptimos para el primal $(\mathbf{x}^*)$ y el dual $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, se cumplen las siguientes condiciones:

$$\begin{aligned} f_i(\mathbf{x}^*) &\leq 0, \quad i = 1, \ldots, m. \\ h_i(\mathbf{x}^*) &= 0, \quad i = 1, \ldots, p. \\ \lambda_i^* &\geq 0, \quad i = 1, \ldots, m. \\ \lambda_i^* f_i(\mathbf{x}^*) &= 0, \quad i = 1, \ldots, m. \end{aligned}$$

$$\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(\mathbf{x}^*) = 0$$

- En cualquier problema de optimización con funciones objetivo y de restricciones diferenciables, para el cual se tenga dualidad fuerte, cualquier par de puntos óptimos del primal y el dual deben satisfacer las condiciones de KKT.

- Suponga ahora que el problema primal es convexo.

- Suponga ahora que el problema primal es convexo.
- Suponga que $(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ satisfacen KKT:

$$\begin{aligned} f_i(\tilde{\mathbf{x}}) &\leq 0, \quad i = 1, \ldots, m. \\ h_i(\tilde{\mathbf{x}}) &= 0, \quad i = 1, \ldots, p. \\ \tilde{\lambda}_i &\geq 0, \quad i = 1, \ldots, m. \\ \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) &= 0, \quad i = 1, \ldots, m. \end{aligned}$$

$$\nabla f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{\mathbf{x}}) = 0$$

- Suponga ahora que el problema primal es convexo.
- Suponga que $(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ satisfacen KKT:

$$
\begin{aligned}
f_i(\tilde{\mathbf{x}}) &\leq 0, \quad i = 1, \dots, m. \\
h_i(\tilde{\mathbf{x}}) &= 0, \quad i = 1, \dots, p. \\
\tilde{\lambda}_i &\geq 0, \quad i = 1, \dots, m. \\
\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) &= 0, \quad i = 1, \dots, m.
\end{aligned}
$$

$$
\nabla f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{\mathbf{x}}) = 0
$$

- $\tilde{\mathbf{x}}$ es factible,

- Suponga ahora que el problema primal es convexo.
- Suponga que $(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ satisfacen KKT:

$$
\begin{aligned}
f_i(\tilde{\mathbf{x}}) \leq 0, &\quad i = 1, \ldots, m. \\
h_i(\tilde{\mathbf{x}}) = 0, &\quad i = 1, \ldots, p. \\
\tilde{\lambda}_i \geq 0, &\quad i = 1, \ldots, m. \\
\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0, &\quad i = 1, \ldots, m.
\end{aligned}
$$

$$\nabla f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{\mathbf{x}}) = 0$$

- $\tilde{\mathbf{x}}$ es factible, $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ es convexo en $\mathbf{x}$,

- Suponga ahora que el problema primal es convexo.
- Suponga que $(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ satisfacen KKT:

$$
\begin{aligned}
f_i(\tilde{\mathbf{x}}) \le 0, &\quad i = 1, \dots, m. \\
h_i(\tilde{\mathbf{x}}) = 0, &\quad i = 1, \dots, p. \\
\tilde{\lambda}_i \ge 0, &\quad i = 1, \dots, m. \\
\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0, &\quad i = 1, \dots, m.
\end{aligned}
$$

$$
\nabla f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{\mathbf{x}}) = 0
$$

- $\tilde{\mathbf{x}}$ es factible, $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ es convexo en $\mathbf{x}$, y $\tilde{\mathbf{x}}$ minimiza $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$.

- Tenemos:

- Tenemos:

$$g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) = L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$$

- Tenemos:

$$\begin{aligned} g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) &= L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \\ &= f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{\mathbf{x}}) \end{aligned}$$

- Tenemos:

$$
\begin{aligned}
g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) &= L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \\
&= f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{\mathbf{x}}) \\
&= f_0(\tilde{\mathbf{x}})
\end{aligned}
$$

- Tenemos:

$$\begin{aligned} g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) &= L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \\ &= f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{\mathbf{x}}) \\ &= f_0(\tilde{\mathbf{x}}) \end{aligned}$$

- Es decir $\tilde{\mathbf{x}}$ y $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ tienen brecha de dualidad cero.

- Tenemos:

$$
\begin{aligned}
g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) &= L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \\
&= f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{\mathbf{x}}) \\
&= f_0(\tilde{\mathbf{x}})
\end{aligned}
$$

- Es decir $\tilde{\mathbf{x}}$ y $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ tienen brecha de dualidad cero.
- $\tilde{\mathbf{x}}$ y $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ son óptimos para el primal y el dual.

- Tenemos:

$$\begin{aligned} g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) &= L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \\ &= f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{\mathbf{x}}) \\ &= f_0(\tilde{\mathbf{x}}) \end{aligned}$$

- Es decir $\tilde{\mathbf{x}}$ y $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ tienen brecha de dualidad cero.
- $\tilde{\mathbf{x}}$ y $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ son óptimos para el primal y el dual.
- En cualquier problema de optimización convexo con funciones objetivo y de restricciones diferenciables, cualquier par de puntos que satisfagan KKT son óptimos para el primal y el dual.

Desigualdad de Markov

Desigualdad de Markov

Teorema

Sea X una variable aleatoria con $X \geq 0$ casi seguramente, y $a > 0$.

Desigualdad de Markov

Teorema

Sea X una variable aleatoria con $X \geq 0$ casi seguramente, y $a > 0$. Entonces:

$$\mathbf{P}\left[X \geq a\right] \leq \frac{\mathbf{E}\left[X\right]}{a}$$

Desigualdad de Markov

Teorema

Sea X una variable aleatoria con $X \geq 0$ casi seguramente, y $a > 0$. Entonces:

$$\mathbf{P}\left[X \geq a\right] \leq \frac{\mathbf{E}\left[X\right]}{a}$$

Demostración.

Desigualdad de Markov

Teorema

Sea X una variable aleatoria con $X \geq 0$ casi seguramente, y $a > 0$. Entonces:

$$\mathbf{P}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}$$

Demostración.

$$\mathbf{P}[X \geq a]$$

Desigualdad de Markov

Teorema

Sea X una variable aleatoria con $X \geq 0$ casi seguramente, y $a > 0$. Entonces:

$$\mathbf{P}\left[X \geq a\right] \leq \frac{\mathbf{E}\left[X\right]}{a}$$

Demostración.

$$\mathbf{P}\left[X \geq a\right] = \mathbf{E}\left[I_{\{X \geq a\}}\right]$$

Desigualdad de Markov

Teorema

Sea X una variable aleatoria con $X \geq 0$ casi seguramente, y $a > 0$. Entonces:

$$\mathbf{P}\left[X \geq a\right] \leq \frac{\mathbf{E}\left[X\right]}{a}$$

Demostración.

$$\mathbf{P}\left[X \geq a\right] = \mathbf{E}\left[I_{\{X \geq a\}}\right] \leq \mathbf{E}\left[\frac{X}{a}\right]$$

Desigualdad de Markov

Teorema

Sea X una variable aleatoria con $X \geq 0$ casi seguramente, y $a > 0$. Entonces:

$$\mathbf{P}\left[X \geq a\right] \leq \frac{\mathbf{E}\left[X\right]}{a}$$

Demostración.

$$\mathbf{P}\left[X \geq a\right] = \mathbf{E}\left[I_{\{X \geq a\}}\right] \leq \mathbf{E}\left[\frac{X}{a}\right] = \frac{\mathbf{E}\left[X\right]}{a}$$

□

Desigualdad de Chebyshev

Teorema

Sea X *una variable aleatoria con* $\mathbf{E}[X] = \mu$, $\mathbf{E}\left[(X-\mu)^2\right] = \sigma^2$.

Desigualdad de Chebyshev

Teorema

Sea X una variable aleatoria con $\mathbf{E}[X] = \mu$, $\mathbf{E}\left[(X-\mu)^2\right] = \sigma^2$.
Entonces $\forall a > 0$

$$\mathbf{P}\left[|X-\mu| \geq a\right] \leq \frac{\sigma^2}{a^2}$$

Desigualdad de Chebyshev

Teorema

Sea X una variable aleatoria con $\mathbf{E}[X] = \mu$, $\mathbf{E}\left[(X-\mu)^2\right] = \sigma^2$. Entonces $\forall a > 0$

$$\mathbf{P}\left[|X-\mu| \geq a\right] \leq \frac{\sigma^2}{a^2}$$

Demostración.

$$\mathbf{P}\left[|X-\mu| \geq a\right]$$

Desigualdad de Chebyshev

Teorema

Sea X *una variable aleatoria con* $\mathbf{E}[X] = \mu$, $\mathbf{E}\left[(X-\mu)^2\right] = \sigma^2$.
Entonces $\forall a > 0$

$$\mathbf{P}[|X-\mu| \geq a] \leq \frac{\sigma^2}{a^2}$$

Demostración.

$$\mathbf{P}[|X-\mu| \geq a] = \mathbf{P}\left[|X-\mu|^2 \geq a^2\right]$$

Desigualdad de Chebyshev

Teorema

Sea X una variable aleatoria con $\mathbf{E}[X] = \mu$, $\mathbf{E}\left[(X-\mu)^2\right] = \sigma^2$.
Entonces $\forall a > 0$

$$\mathbf{P}[|X-\mu| \geq a] \leq \frac{\sigma^2}{a^2}$$

Demostración.

$$\mathbf{P}[|X-\mu| \geq a] = \mathbf{P}\left[|X-\mu|^2 \geq a^2\right] \underset{\substack{\downarrow \\ \text{Markov}}}{\leq} \frac{\sigma^2}{a^2}$$

$\square$

Desigualdad de Hoeffding

Teorema

Sean $X_1, X_2, \ldots, X_n$ *variables aleatorias independientes, con*

Desigualdad de Hoeffding

Teorema

Sean $X_1, X_2, \ldots, X_n$ *variables aleatorias independientes, con*

- $\mathbf{E}[X_j] = 0$ *para* $j = 1, \ldots, n$.

Desigualdad de Hoeffding

Teorema

Sean $X_1, X_2, \ldots, X_n$ *variables aleatorias independientes, con*

- $\mathbf{E}[X_j] = 0$ *para* $j = 1, \ldots, n$.
- $a_j \leq X_j \leq b_j$, *con* $a_j, b_j \in \mathbb{R}$ *para* $j = 1, \ldots, n$.

Desigualdad de Hoeffding

Teorema

Sean $X_1, X_2, \ldots, X_n$ *variables aleatorias independientes, con*

- $\mathbf{E}[X_j] = 0$ *para* $j = 1, \ldots, n$.
- $a_j \leq X_j \leq b_j$, *con* $a_j, b_j \in \mathbb{R}$ *para* $j = 1, \ldots, n$.

entonces:

$$\mathbf{P}\left[\sum_{j=1}^{n} X_j \geq \epsilon\right] \leq \exp\left(-\frac{2\epsilon^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\right)$$

$$\mathbf{P}\left[\sum_{j=1}^{n} X_j \leq -\epsilon\right] \leq \exp\left(-\frac{2\epsilon^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\right)$$

Desigualdad de Hoeffding

Teorema

Sean $X_1, X_2, \ldots, X_n$ *variables aleatorias independientes, con*

- $\mathbf{E}[X_j] = 0$ *para* $j = 1, \ldots, n$.
- $a_j \le X_j \le b_j$, *con* $a_j, b_j \in \mathbb{R}$ *para* $j = 1, \ldots, n$.

entonces:

$$\mathbf{P}\left[\sum_{j=1}^{n} X_j \ge \epsilon\right] \le \exp\left(-\frac{2\epsilon^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\right)$$

$$\mathbf{P}\left[\sum_{j=1}^{n} X_j \le -\epsilon\right] \le \exp\left(-\frac{2\epsilon^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\right)$$

o combinando:

$$\mathbf{P}\left[\left|\sum_{j=1}^{n} X_j\right| \ge \epsilon\right] \le 2\exp\left(-\frac{2\epsilon^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\right)$$

Caso particular: Cotas de Chernoff

Caso particular: Cotas de Chernoff

- $X_j \in \{0, 1\}$, $\mathbf{P}[X_j = 1] = p$

Caso particular: Cotas de Chernoff

- $X_j \in \{0,1\}$, $\mathbf{P}[X_j = 1] = p$
- $\frac{(X_j - p)}{n} \in$

Caso particular: Cotas de Chernoff

- $X_j \in \{0,1\}$, $\mathbf{P}\left[X_j = 1\right] = p$
- $\frac{(X_j - p)}{n} \in \left\{-\frac{p}{n}, \frac{1-p}{n}\right\}$

Caso particular: Cotas de Chernoff

- $X_j \in \{0, 1\}$, $\mathbf{P}[X_j = 1] = p$
- $\frac{(X_j - p)}{n} \in \left\{-\frac{p}{n}, \frac{1-p}{n}\right\} \Rightarrow b_j - a_j =$

Caso particular: Cotas de Chernoff

- $X_j \in \{0, 1\}$, $\mathbf{P}[X_j = 1] = p$
- $\frac{(X_j - p)}{n} \in \left\{ -\frac{p}{n}, \frac{1-p}{n} \right\} \Rightarrow b_j - a_j = \frac{1}{n}$

Caso particular: Cotas de Chernoff

- $X_j \in \{0, 1\}$, $\mathbf{P}[X_j = 1] = p$
- $\frac{(X_j - p)}{n} \in \left\{-\frac{p}{n}, \frac{1-p}{n}\right\} \Rightarrow b_j - a_j = \frac{1}{n}$
- Por Hoeffding:

Caso particular: Cotas de Chernoff

- $X_j \in \{0,1\}$, $\mathbf{P}[X_j = 1] = p$
- $\frac{(X_j - p)}{n} \in \left\{-\frac{p}{n}, \frac{1-p}{n}\right\} \Rightarrow b_j - a_j = \frac{1}{n}$
- Por Hoeffding:

$$\mathbf{P}\left[\frac{1}{n}\sum_{j=1}^{n} X_j - p \geq \varepsilon\right] \leq e^{-2\epsilon^2 n}$$

y

$$\mathbf{P}\left[\frac{1}{n}\sum_{j=1}^{n} X_j - p \leq -\varepsilon\right] \leq e^{-2\epsilon^2 n}$$

Caso particular: Cotas de Chernoff

- $X_j \in \{0,1\}$, $\mathbf{P}[X_j = 1] = p$
- $\frac{(X_j - p)}{n} \in \left\{-\frac{p}{n}, \frac{1-p}{n}\right\} \Rightarrow b_j - a_j = \frac{1}{n}$
- Por Hoeffding:

$$\mathbf{P}\left[\frac{1}{n}\sum_{j=1}^{n} X_j - p \geq \varepsilon\right] \leq e^{-2\epsilon^2 n}$$

y

$$\mathbf{P}\left[\frac{1}{n}\sum_{j=1}^{n} X_j - p \leq -\varepsilon\right] \leq e^{-2\epsilon^2 n}$$

- Estas son las cotas de Chernoff en forma aditiva.

Ejemplo

Ejemplo

- Moneda, estimar probabilidad p de que salga cara.

Ejemplo

- Moneda, estimar probabilidad p de que salga cara.
- Estimativo $\hat{p} =$ número de caras en n lanzadas.

Ejemplo

- Moneda, estimar probabilidad p de que salga cara.
- Estimativo $\hat{p} =$ número de caras en n lanzadas.
- Cuántas veces tenemos que lanzar la moneda para garantizar con confianza $1 - \delta$ que el estimativo $\hat{p}$ no difiera en más de ϵ de p?

Ejemplo

- Moneda, estimar probabilidad p de que salga cara.
- Estimativo $\hat{p} =$ número de caras en n lanzadas.
- Cuántas veces tenemos que lanzar la moneda para garantizar con confianza $1 - \delta$ que el estimativo $\hat{p}$ no difiera en más de ϵ de p?
- Usando cotas de Chernoff:

$$\mathbf{P}\left[|p - \hat{p}| \geq \epsilon\right] \leq 2e^{-2\epsilon^2 n}$$

Ejemplo

- Moneda, estimar probabilidad p de que salga cara.
- Estimativo $\hat{p} =$ número de caras en n lanzadas.
- Cuántas veces tenemos que lanzar la moneda para garantizar con confianza $1-\delta$ que el estimativo $\hat{p}$ no difiera en más de ϵ de p?
- Usando cotas de Chernoff:

$$\mathbf{P}\left[|p-\hat{p}| \geq \epsilon\right] \leq 2e^{-2\epsilon^2 n}$$

Queremos $2e^{-2\epsilon^2 n} = \delta$

Ejemplo

- Moneda, estimar probabilidad p de que salga cara.
- Estimativo $\hat{p}$ = número de caras en n lanzadas.
- Cuántas veces tenemos que lanzar la moneda para garantizar con confianza $1-\delta$ que el estimativo $\hat{p}$ no difiera en más de ϵ de p?
- Usando cotas de Chernoff:

$$\mathbf{P}\left[|p-\hat{p}| \geq \epsilon\right] \leq 2e^{-2\epsilon^2 n}$$

Queremos $2e^{-2\epsilon^2 n} = \delta$ o despejando $n = \frac{1}{2\epsilon^2} \ln \frac{2}{\delta}$.

Ejemplo

- Moneda, estimar probabilidad p de que salga cara.
- Estimativo $\hat{p}$ = número de caras en n lanzadas.
- Cuántas veces tenemos que lanzar la moneda para garantizar con confianza $1-\delta$ que el estimativo $\hat{p}$ no difiera en más de ϵ de p?
- Usando cotas de Chernoff:

$$\mathbf{P}\left[|p-\hat{p}| \geq \epsilon\right] \leq 2e^{-2\epsilon^2 n}$$

  Queremos $2e^{-2\epsilon^2 n} = \delta$ o despejando $n = \frac{1}{2\epsilon^2} \ln \frac{2}{\delta}$.
- Por ejemplo para confianza del 95 % y precisión 0,05 debemos lanzar la moneda ~ 738 veces.