Lagrange Multipliers

Constrained optimization
Constrained optimization

- What constraints might we want for ML?
  - Probabilities sum to 1
  - Regression weights non-negative
  - Regression weights less than a constant

- More generally
  - Fixed amount of money or time or energy available
To maximize $f(x, y)$ subject to $g(x, y) = k$

find:

- The largest value of $c$ such that the level curve $f(x, y) = c$ intersects $g(x, y) = k$.
- This happens when the lines are parallel

$$\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)$$
Lagrange Multiplier – the idea

Find

\[ \min_x f(x) \]  \hspace{1cm} -- x was (x,y) on the last slide

s.t.

\[ c_i(x) = 0 \quad j=1\ldots m \]  \hspace{1cm} -- c_1(x) was g(x)-k on the last slide

Set

\[ L(x, \lambda) = f(x) + \lambda^T c(x) \]

At the minimum of \( L(x, \lambda) \)

\[ \frac{dL}{dx} = \frac{df}{dx} + \lambda^T \frac{dc}{dx} = 0 \]

\[ \frac{dL}{d\lambda} = c(x) = 0 \]

This makes the curves be parallel
As on the last slide
Lagrange Multiplier – generalization

Find

\[ \min \limits_x f(x) \]

s.t.

\[ c_i(x) \leq 0 \quad j=1\ldots m \]

Set

\[ L(x,\lambda) = f(x) + \lambda^T c(x) \]

At the minimum of \( L(x,\lambda) \)

\[ \frac{dL}{dx} = \frac{df}{dx} + \lambda^T \frac{dc}{dx} = 0 \]

\[ \lambda_i c_i(x) = 0 \quad j=1\ldots m \]

\[ \lambda_i \geq 0 \quad j=1\ldots m \]

KKT = Karush Kuhn Tucker conditions

For each \( \lambda_j \), either

\[ \lambda_j = 0 \] (the constraint is not active)

or

\[ \lambda_j > 0 \] (the constraint is active)

and thus

\[ c_i(x) = 0 \]
Lagrange Multiplier Steps

1. Start with the primal
   \[
   \text{minimize } f_0(x), \\
   \text{subject to } f_i(x) \leq 0, \quad i = 1, \ldots, m \\
   h_i(x) = 0, \quad i = 1, \ldots, p
   \]

2. Formulate \( L \)
   \[
   L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x).
   \]

3. Find \( g(\lambda) = \min_x (L) \)
   solve \( dL/dx = 0 \)

4. Find \( \max g(\lambda, \nu) \) s.t. \( \lambda_i \geq 0 \quad \nu_i \geq 0 \)

5. See if the constraints are binding

6. Find \( x^* \)
   \[
   g(\lambda^*) = f(x^*).
   \]
Lagrange Multiplier Steps

1. Start with the primal

\[
\text{minimize } \frac{1}{2}cx^2 \quad \text{subject to } \quad ax - b \leq 0
\]

2. Formulate \( L \)

\[
L(x, \lambda) = \frac{1}{2}cx^2 + \lambda(ax - b).
\]

3. Find \( g(\lambda) = \min_x (L) \)

solve \( \frac{dL}{dx} = 0 \)

\[
\begin{align*}
\text{plug back into } L \\
\end{align*}
\]

\[
\begin{align*}
\frac{dL}{dx} &= dx + \lambda a \\
&= 0 \\
x &= -\lambda \frac{a}{c}. \\
g(\lambda) &= \frac{1}{2}c \left(-\lambda \frac{a}{c}\right)^2 + \lambda a \left(-\lambda \frac{a}{c}\right) - \lambda b \\
&= -\frac{a^2}{2c} \lambda^2 - \lambda b.
\end{align*}
\]

4. Find \( \max g(\lambda, \nu) \) s.t. \( \lambda_i \geq 0 \)

try maximizing without constraints

\[
\frac{-a^2}{c} \lambda - b = 0 \quad \Rightarrow \quad \lambda^* = \frac{-bc}{a^2}.
\]

5. See if the constraints are binding

it depends on the sign of \(-bc\)

6. Find \( x^* \)

plug \( \lambda^* \) into relation

\[
x = -\lambda \frac{a}{c} \quad = \frac{b}{a}
\]
Lagrange Multipliers Visually

\[ \text{min } \frac{1}{2} x^2 \]
\[ \text{s.t. } 2x + 5 > 0 \]

feasible

infeasible

\[ a=2, \ b=-5, \ c=1 \]
Solve

maximize

\[ f(x,y) = x + y \]

subject to

\[ x^2 + y^2 - 1 = 0 \]

Answer: \( x^* = (x,y) = ?? \)

1. Formulate \( L = f_0(x) + \lambda f_1(x) \)
2. Find \( \min_x (L) = g(\lambda) \)
3. Find \( \max g(\lambda) \)
4. See if constraints are binding
5. Find \( x^* \)

Note that we formulate the problem in terms of minimization!!!
The answer
Formulate and solve

Find values of a set of $k$ probabilities $(p_1, p_2, \ldots p_k)$ that maximize their entropy

minimize

$f(p) = ??$

subject to

??

Answer: $p_i = ??$

1. Formulate $L$
2. Find $\min_x (L) = g(\lambda)$
3. Find $\max g(\lambda)$
4. See if constraints are binding
5. Find $x^*$
Formulate

Find values of a vector of $p$ non-negative weights $w$ that minimize $\sum_i (y_i - x_i w)^2$

minimize

$f(w) = ??$

subject to

??