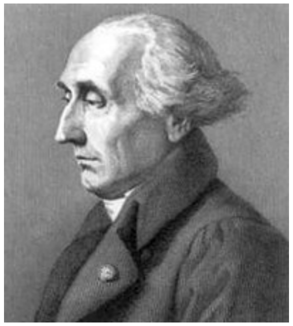
Constrained optimization

Lagrange Multipliers

Les mathématiques sont comme le porc, tout en est bon.
-Joseph-Louis Lagrange



Constrained optimization

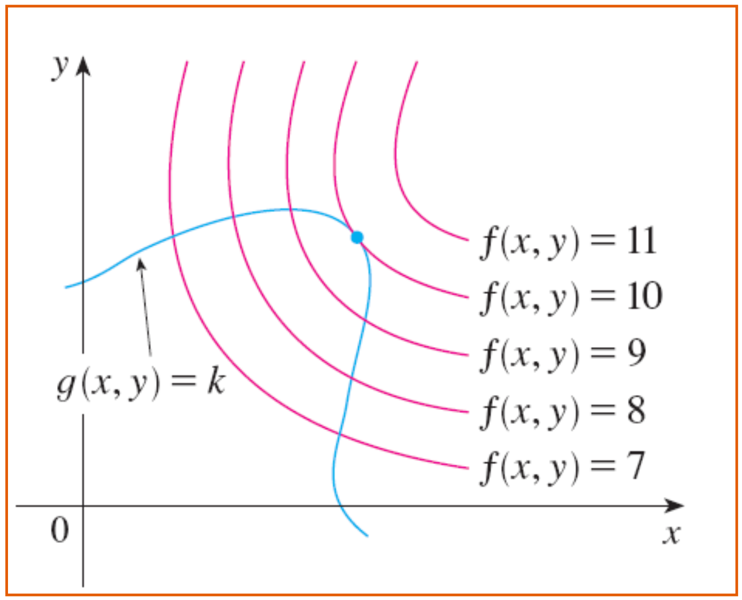
- **What constraints might we want for ML?**
  - Probabilities sum to 1
  - Regression weights non-negative
  - Regression weights less than a constant
- **More generally**
  - Fixed amount of money or time or energy available

Lagrange multipliers - example

- **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

$$\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)$$

This example *maximizes*; we'll mostly *minimize*



www2.essex.edu/~wang/221/Chap14_Sec8.ppt

The same approach works for inequality constraints

Lagrange multipliers – the idea

Find

min$_x$ $f(\mathbf{x})$ -- $\mathbf{x}$ *was (x,y) on the last slide*

s.t.

$c_i(\mathbf{x}) = 0 \quad j=1\ldots m$ -- $c_1(x)$ *was g(x)-k on the last slide*

Set

$L(\mathbf{x},\lambda) = f(\mathbf{x}) + \lambda^T \mathbf{c}(\mathbf{x})$

At the minimum of $L(\mathbf{x},\lambda)$

$dL/d\mathbf{x} = df/d\mathbf{x} + \lambda^T d\mathbf{c}/d\mathbf{x} = 0$

$dL/d\lambda = \mathbf{c}(\mathbf{x}) = 0$

This makes the curves be parallel
As on the last slide

Lagrange multipliers – generalization

Find

$\min_x f(\boldsymbol{x})$

s.t.

$c_i(\boldsymbol{x}) \leq 0 \quad j=1\ldots m$

Set

$L(\mathbf{x},\lambda) = f(\mathbf{x}) + \lambda^T \mathbf{c}(\mathbf{x})$

At the minimum of $L(\mathbf{x},\lambda)$

$dL/d\mathbf{x} = df/d\mathbf{x} + \lambda^T d\mathbf{c}/d\mathbf{x} = 0$

$\lambda_i c_i(\mathbf{x}) = 0 \quad j=1\ldots m$

$\lambda_i \geq 0 \quad j=1\ldots m$

KKT = Karush Kuhn Tucker conditions

Complementary Slackness

For each λ_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**

$$\begin{array}{ll} \text{minimize} & f_0(x), \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{array}$$

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. $\lambda_i >= 0 \quad \nu_i >= 0$

5. **See if the constraints are binding**

6. **Find** x^*

$$g(\lambda^\star) = f(x^\star)$$

Lagrange multiplier steps

1. Start with the primal

$$\text{minimize} \quad \tfrac{1}{2}cx^2 \quad \text{subject to} \quad ax - b \le 0$$

2. Formulate *L*

$$L(x,\lambda) = \tfrac{1}{2}cx^2 + \lambda(ax - b).$$

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

solve $dL/dx = 0$

$$cx + \lambda a = 0 \quad \rightarrow \quad x = -\lambda \tfrac{a}{c}.$$

plug back into L

$$\begin{aligned} g(\lambda) &= \tfrac{1}{2}c\left(-\lambda\tfrac{a}{c}\right)^2 + \lambda a\left(-\lambda\tfrac{a}{c}\right) - \lambda b \\ &= \tfrac{-a^2}{2c}\lambda^2 - \lambda b. \end{aligned}$$

4. Find $\max g(\lambda, \nu)$ s.t. $\lambda_i >= 0$

try maximizing without constraints

$$\tfrac{-a^2}{c}\lambda - b = 0 \quad \rightarrow \quad \lambda^* = \tfrac{-bc}{a^2}.$$

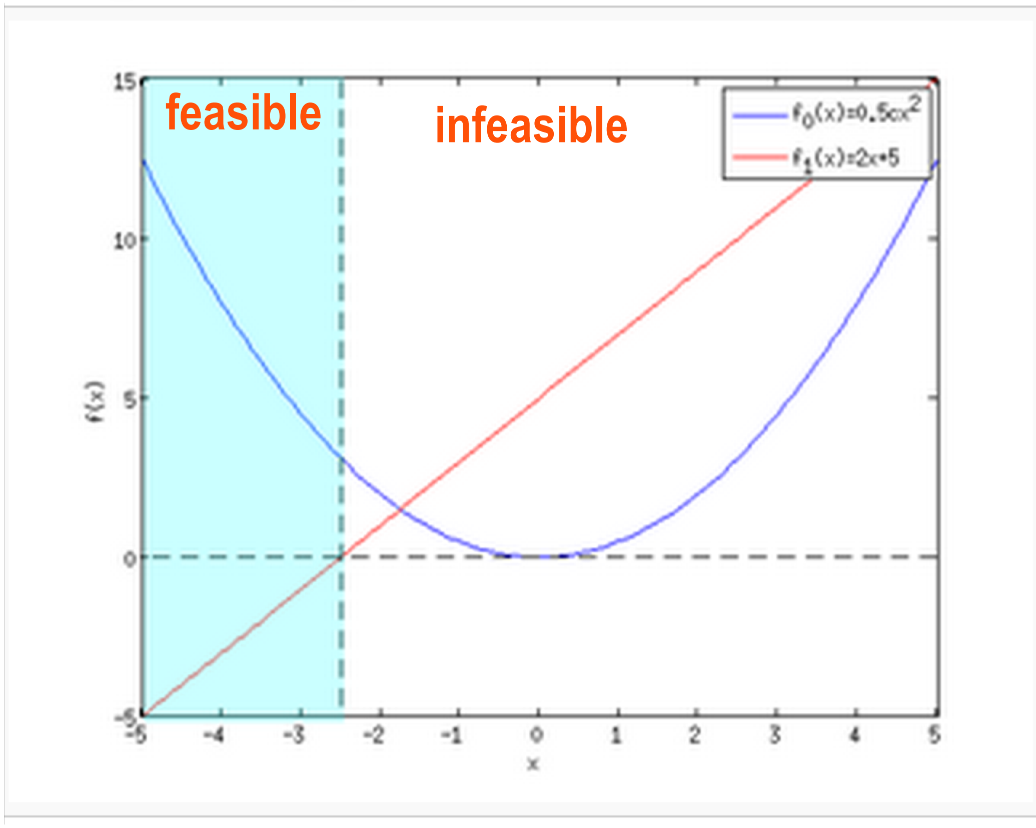
5. See if the constraints are binding

it depends on the sign of –bc

6. Find x^*

plug λ^* into x= - λ (a/c) giving x = *b/a*

Lagrange multipliers visually



min (1/2) x^2
s.t.
$2x + 5 < 0$

a=2, b=-5, c=1

Solve

maximize

$$f(x,y) = x + y$$

subject to

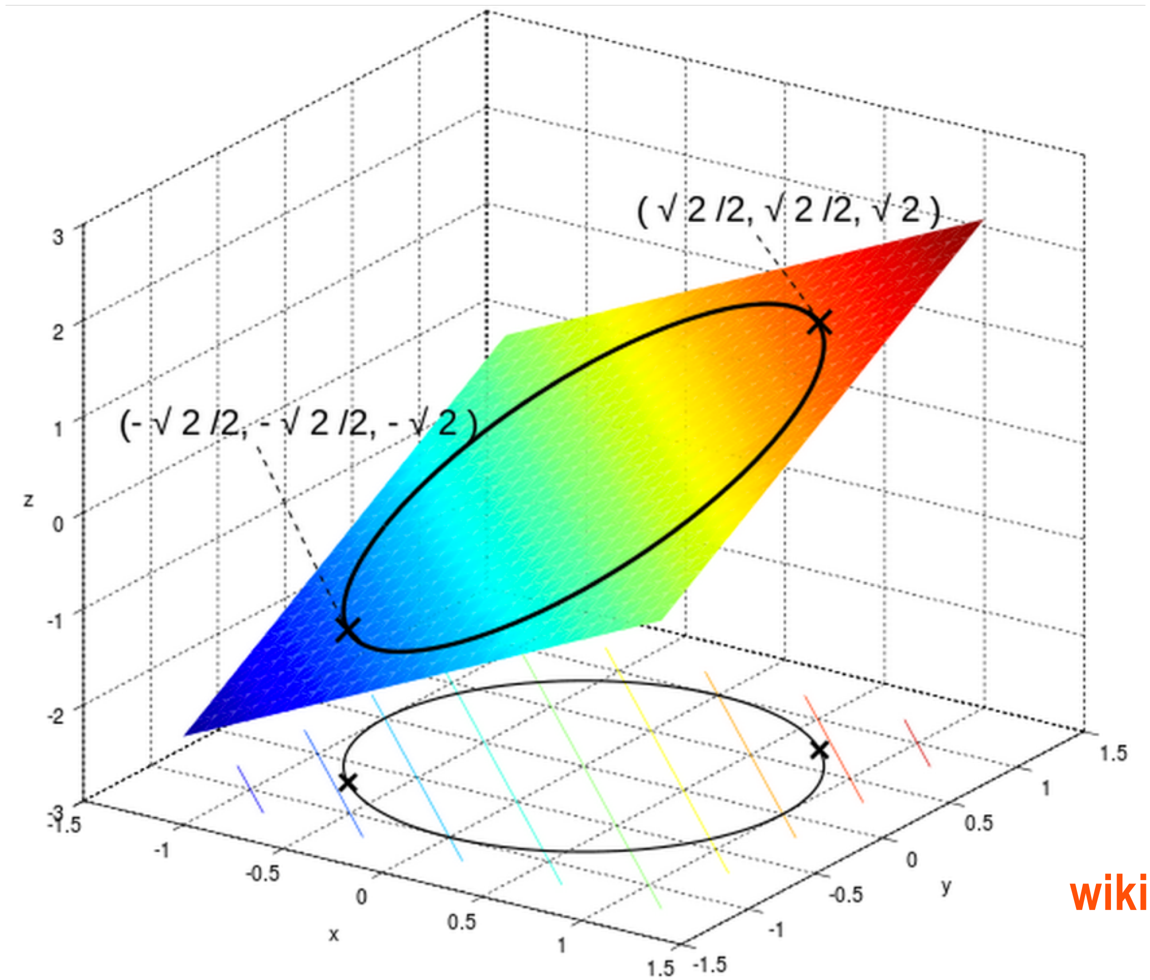
$$x^2 + y^2 - 1 = 0$$

Answer: $\mathbf{x}^* = (x,y) = ??$

1. Start with the primal
2. Formulate $L = f_0(\mathbf{x}) + \lambda\, f_1(\mathbf{x})$
3. Find $\min_{\mathbf{x}} (L) = g(\lambda)$
4. Find *max* $g(\lambda)$
5. See if constraints are binding
6. Find $\mathbf{x}^*$

Note that we formulate the problem in terms of minimization!!!

The answer



wikipedia

Formulate and solve

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

1. Start with the primal
2. Formulate L
3. Find $\min_{\mathbf{x}} (L) = g(\lambda)$
4. Find $\max g(\lambda)$
5. See if constraints are binding
6. Find x^*

minimize

$\boldsymbol{f(p) = \;??}$

subject to

??

Answer: $\boldsymbol{p_i = ??}$

Formulate

Find values of a vector of *p* non-negative weights *w* that minimize $\Sigma_i(y_i - \mathbf{x}_i^T w)^2$

minimize

$f(w) = ??$

subject to

??

Lagrange multiplier steps

1. Start with the primal

   $min_w \, \Sigma_i(y_i\textbf{-}\mathbf{x}_i\mathbf{w})^2$

   $s.t. \; - w_j < 0$

2. Formulate L

   $L(\mathbf{w}, \lambda) = \Sigma_i(y_i\textbf{-}\mathbf{x}_i\mathbf{w})^2 \; - \; \Sigma_j \lambda_j w_j \; = \Sigma_i(y_i\textbf{-}\mathbf{x}_i\mathbf{w})^2 \; - \; \boldsymbol{\lambda}^T \mathbf{w}$

3. Find $g(\lambda) = \; min_{\mathbf{w}} (L)$

   solve $dL/d\mathbf{w} = 0 \quad 2\Sigma_i \mathbf{x}_i^T(y_i\textbf{-}\mathbf{x}_i\mathbf{w}) - \; diag(\lambda) = 0$

   $\mathbf{w} = (\mathbf{X'X} + diag(\lambda)/2)^{-1} \mathbf{X'y}$

   plug back into L

   $g(\lambda) = \Sigma_i(y_i\textbf{-}\mathbf{x}_i(\mathbf{X'X} + diag(\lambda)/2)^{-1} \mathbf{X'y})^2 \; -$
   $\boldsymbol{\lambda}^T (\mathbf{X'X} + diag(\lambda)/2)^{-1} \mathbf{X'y}$

4. Find $max \; g(\lambda,\nu) \;\; s.t. \; \lambda_\iota >= 0$

   try maximizing without constraints

5. See if the constraints are binding
6. Find x^*

   plug λ^* into relation

Lagrange solution properties

- **Weak duality:** The optimal value of the dual problem *always* gives a lower bound on the optimal value of the primal problem
- **Strong duality:** *If the primal problem is convex*,
  - *i.e.* f and g_i are convex functions and h_j are affine (i.e. linear) functions

  and some fine print math is true

  then the *primal and dual optimal values are equal*

What you should know

- **The steps to solve (simple) constrained optimization problems**
- **Complementary slackness**
  - Either $\lambda_j = 0$ and the constraint is *not binding* (not active)
  - Or $\lambda_j > 0$ and the constraint is *binding* (active)
- **Concept of weak and strong duality**