Lagrange Multipliers

Constrained optimization

Lyle Ungar, University of Pennsylvania

Constrained optimization

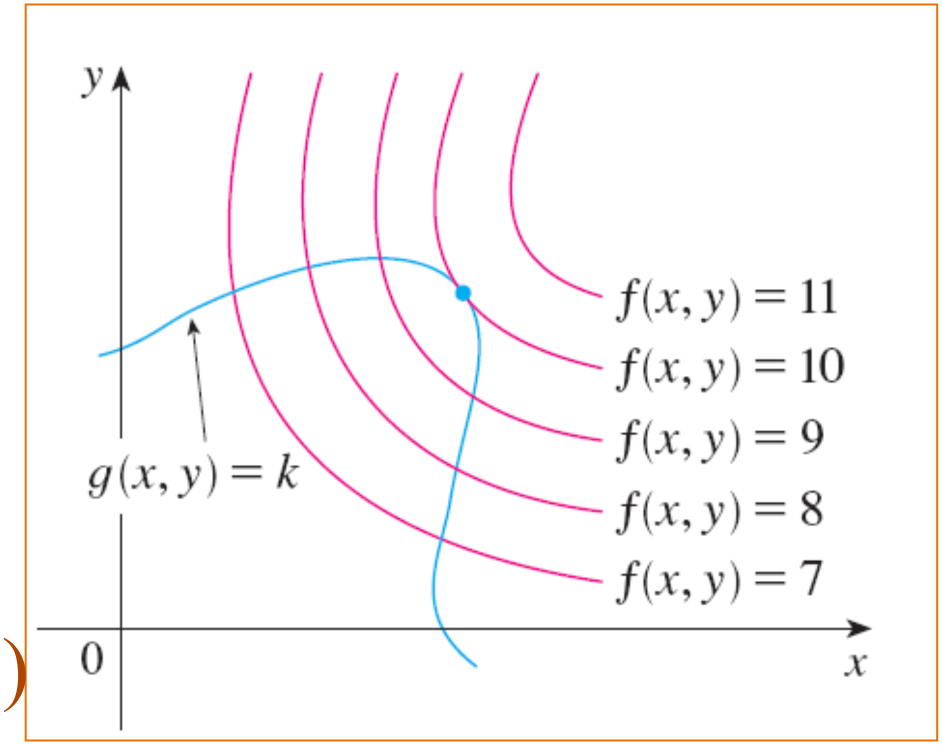
- **What constraints might we want for ML?**
  - Probabilities sum to 1
  - Regression weights non-negative
  - Regression weights less than a constant
- **More generally**
  - Fixed amount of money or time or energy available

Lagrange Mulipliers - 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

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



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

Lagrange Multiplier – the idea

Find

$\min_x f(\boldsymbol{x})$ -- $\boldsymbol{x}$ *was (x,y) on the last slide*

s.t.

$c_i(\boldsymbol{x}) = 0 \quad j=1\ldots m$ -- *$c_1(x)$ 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 Multiplier – generalization

Find

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

s.t.

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

Set

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

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

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

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

$\boldsymbol{\lambda}_i \ge 0 \quad j=1\ldots m$

For each λ_j, either

$\lambda_j = 0$ (the constraint is not active)

or

$\lambda_j > 0$ (the constraint is active)

and thus

$c_i(x) = 0$

KKT = Karush Kuhn Tucker conditions

Lagrange Multiplier Steps

1. Start with the primal

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

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

5. See if the constraints are binding

6. Find x^*

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

Lagrange Multiplier Steps

1. Start with the primal

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

2. Formulate *L*

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

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

   solve $dL/dx = 0$ $\quad cx + \lambda a = 0 \quad \to \quad x = -\lambda \frac{a}{c}.$

   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

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

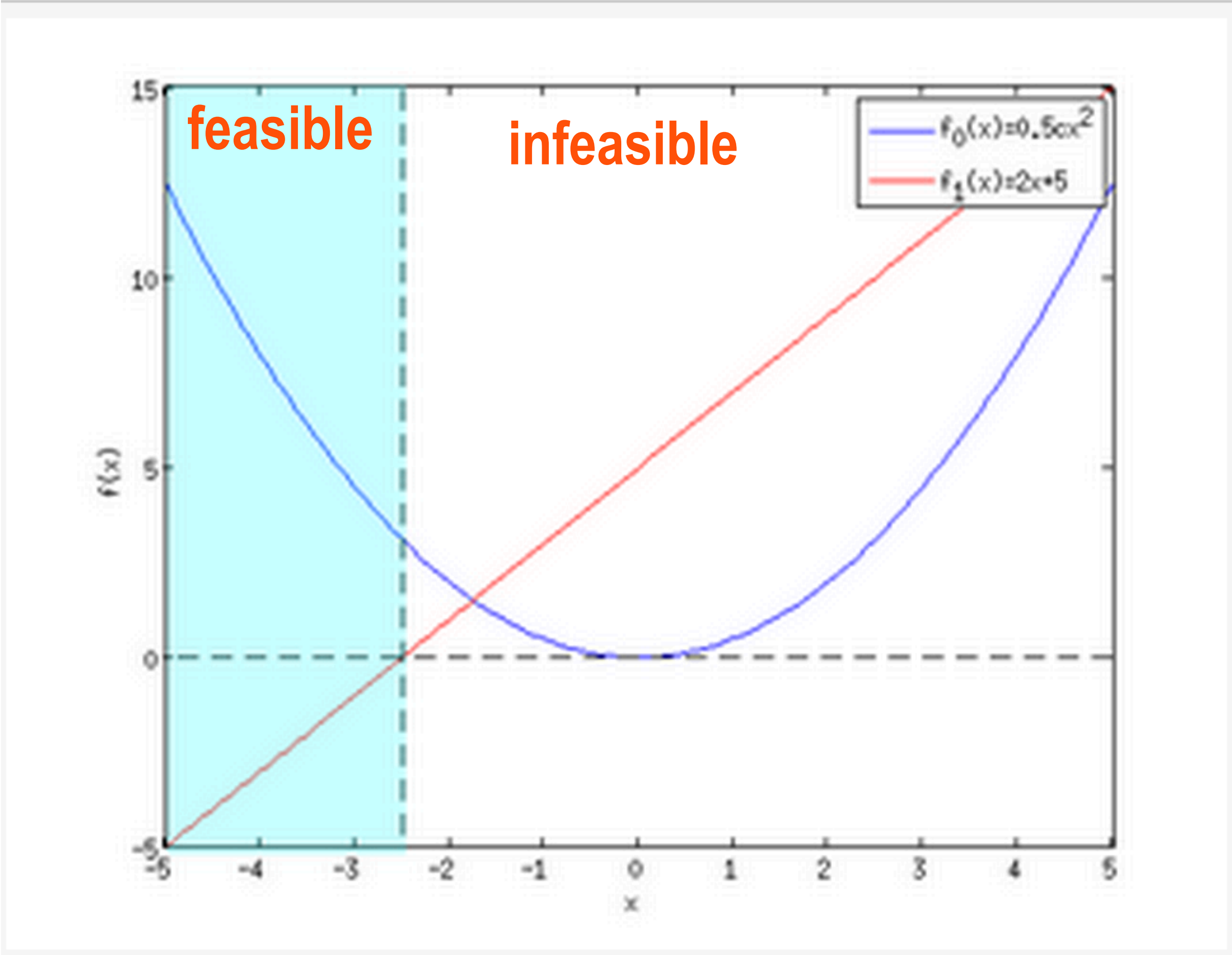
5. See if the constraints are binding

   it depends on the sign of –bc

6. Find x^*

   plug λ^* into relation $x = -\lambda\frac{a}{c}.$ $= b/a$

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

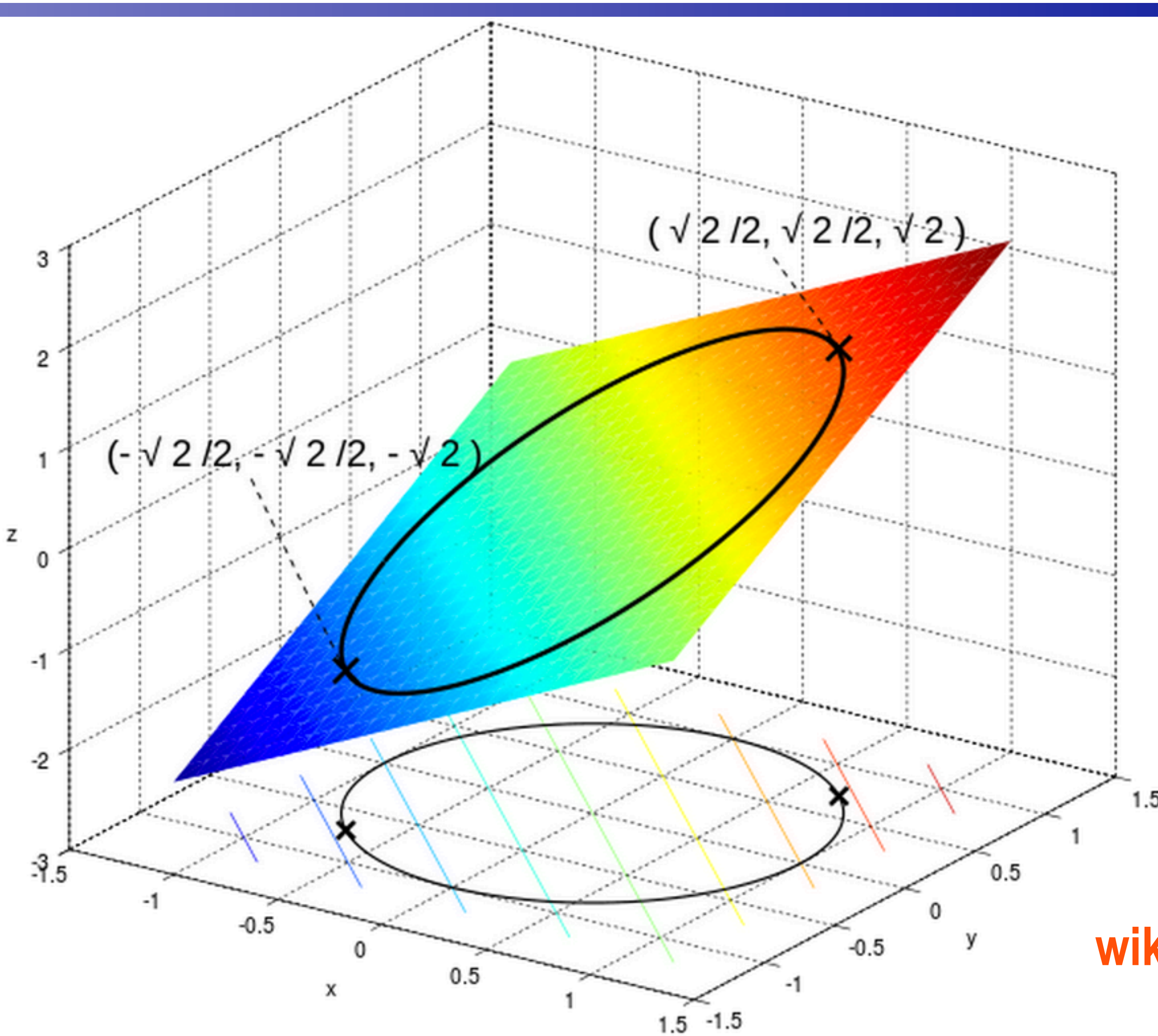
$$x^2 + y^2 - 1 = 0$$

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

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

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

minimize

$f(w) = \; ??$

subject to

??