

CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019



Outline for March 25

- Recap SVM introduction and idea of SVMs as an optimization problem
 - Detour to Lagrange multipliers as a way to solve optimization problems
 - Reformulation of SVMs using inner products between examples
- Office hours TODAY: 12:30-2pm
 - Lab 5 due Tuesday night
 - Fill out partner form for Lab 6 (one week lab)

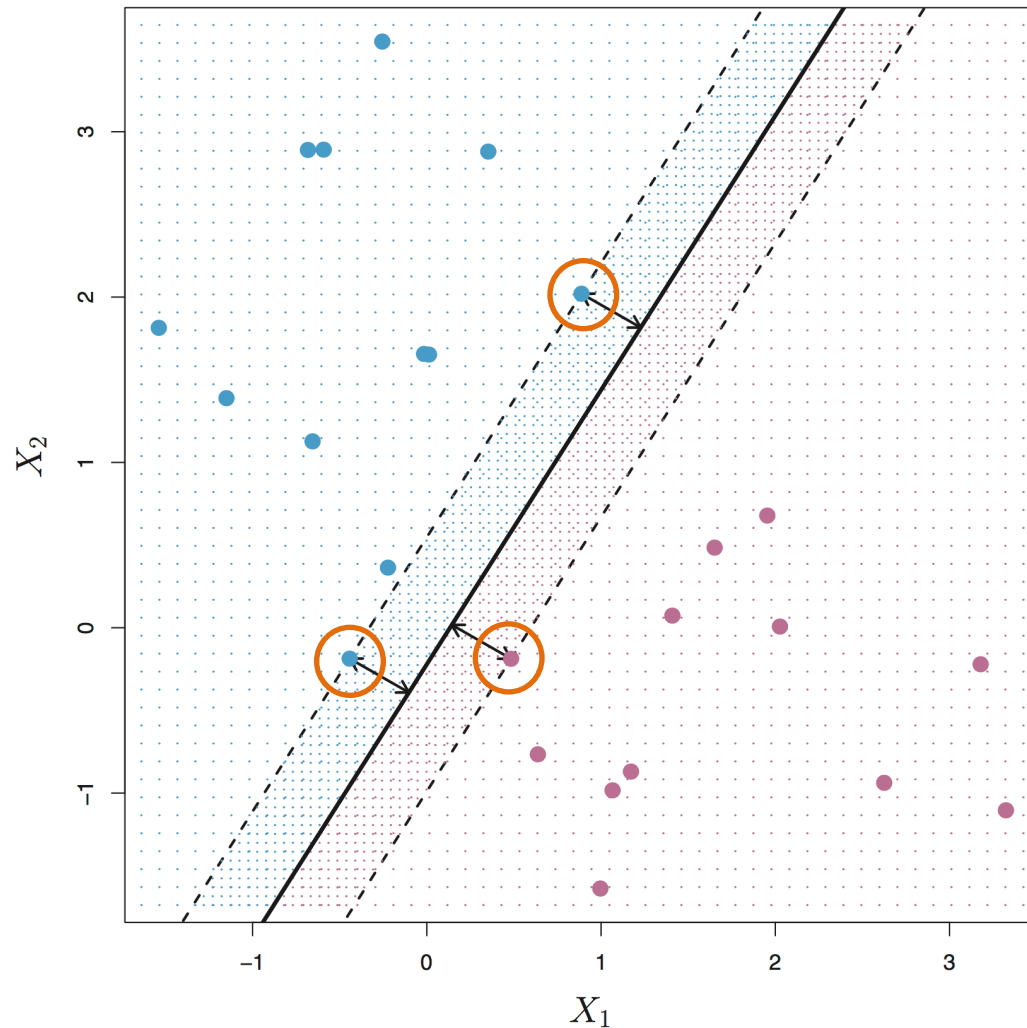
Outline for March 25

- Recap SVM introduction and idea of SVMs as an optimization problem
- Detour to Lagrange multipliers as a way to solve optimization problems
- Reformulation of SVMs using inner products between examples

Recap SVMs

- 1) Won't just find any separating hyperplane, but the “best” one (largest margin)
- 2) Will be able to accommodate data that is not linearly separable
- 3) Will be able to accommodate non-linear boundaries

Goal: find the maximum margin classifier
Datapoints on the margin are “support vectors”



Support vectors

Functional and Geometric Margins

SVM classifier:
(same as perceptron)

$$h(\vec{x}) = \text{sign}(\vec{w} \cdot \vec{x} + b)$$

Functional and Geometric Margins

SVM classifier:
(same as perceptron)

$$h(\vec{x}) = \text{sign}(\vec{w} \cdot \vec{x} + b)$$

Functional Margin:

$$\hat{\gamma}_i = y_i(\vec{w} \cdot \vec{x}_i + b)$$

Functional and Geometric Margins

SVM classifier:
(same as perceptron)

$$h(\vec{x}) = \text{sign}(\vec{w} \cdot \vec{x} + b)$$

Functional Margin:

$$\hat{\gamma}_i = y_i(\vec{w} \cdot \vec{x}_i + b)$$

Geometric Margin:
(distance between
example and hyperplane)

$$\gamma_i = y_i \left(\frac{\vec{w}}{\|\vec{w}\|} \cdot \vec{x}_i + \frac{b}{\|\vec{w}\|} \right)$$

Functional and Geometric Margins

SVM classifier:
(same as perceptron)

$$h(\vec{x}) = \text{sign}(\vec{w} \cdot \vec{x} + b)$$

Functional Margin:

$$\hat{\gamma}_i = y_i(\vec{w} \cdot \vec{x}_i + b)$$

Geometric Margin:
(distance between
example and hyperplane)

$$\gamma_i = y_i \left(\frac{\vec{w}}{\|\vec{w}\|} \cdot \vec{x}_i + \frac{b}{\|\vec{w}\|} \right)$$

Note:

$$\gamma_i = \frac{\hat{\gamma}_i}{\|\vec{w}\|}$$

Optimization Problem: try 1

Goal: maximize the minimum distance
between example and hyperplane

$$\gamma = \min_{i=1, \dots, n} \gamma_i$$

Optimization Problem: try 1

Goal: maximize the minimum distance
between example and hyperplane

$$\gamma = \min_{i=1, \dots, n} \gamma_i$$

Formulation: optimize a function with
respect to a constraint

$$\max_{\gamma, \vec{w}, b} \quad \gamma$$

$$\text{s.t.} \quad y_i(\vec{w} \cdot \vec{x}_i + b) \geq \gamma, \quad i = 1, \dots, n$$

$$\text{and} \quad \|\vec{w}\| = 1$$

(force functional and geometric
margin to be equal)

Optimization Problem: try 2

Idea: substitute functional margin
divided by magnitude of weight vector

$$\begin{aligned} \max_{\hat{\gamma}, \vec{w}, b} \quad & \frac{\hat{\gamma}}{\|\vec{w}\|} \\ \text{s.t.} \quad & y_i(\vec{w} \cdot \vec{x}_i + b) \geq \hat{\gamma}, \quad i = 1, \dots, n \end{aligned}$$

(gets rid of non-convex constraint)

Optimization Problem: try 3

Idea: put arbitrary constraint on functional margin

$$\hat{\gamma} = 1$$

$$\begin{array}{ll} \min_{\vec{w}, b} & \frac{1}{2} \|\vec{w}\|^2 \\ \text{s.t.} & y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1, \quad i = 1, \dots, n \end{array}$$

Optimization Problem: try 3

Idea: put arbitrary constraint on functional margin

$$\hat{\gamma} = 1$$

$$\begin{aligned} \min_{\vec{w}, b} \quad & \frac{1}{2} \|\vec{w}\|^2 \\ \text{s.t.} \quad & y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1, \quad i = 1, \dots, n \end{aligned}$$

$$\begin{aligned} \min_{\vec{w}, b} \quad & \frac{1}{2} \|\vec{w}\|^2 \\ \text{s.t.} \quad & -y_i(\vec{w} \cdot \vec{x}_i + b) + 1 \leq 0, \quad i = 1, \dots, n \end{aligned}$$

Outline for March 25

- Recap SVM introduction and idea of SVMs as an optimization problem
- Detour to Lagrange multipliers as a way to solve optimization problems
- Reformulation of SVMs using inner products between examples

Lagrange Multipliers

$$\max_{x,y} f(x,y)$$

$$\text{s.t. } g(x,y) = 0$$

$$L(x,y,\lambda) = f(x,y) - \lambda \cdot g(x,y)$$

Lagrangian

Lagrange multiplier

require

$$\nabla L(x,y,\lambda) = \vec{0}$$

gradient

$$= \begin{bmatrix} \frac{\partial L}{\partial x} \\ \frac{\partial L}{\partial y} \\ \frac{\partial L}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\frac{\partial h}{\partial \lambda} = \boxed{g(x, y) = 0} \quad \checkmark \quad 1 \text{ equation}$$

$$\Rightarrow \boxed{\nabla f(x, y) = \lambda \cdot \nabla g(x, y)}$$

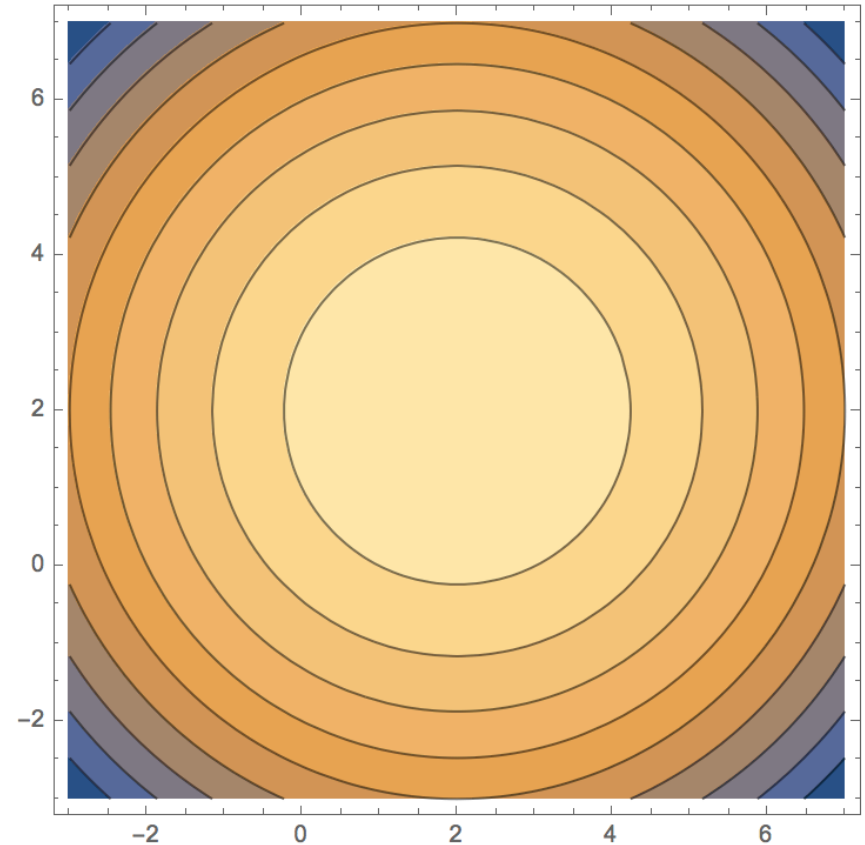
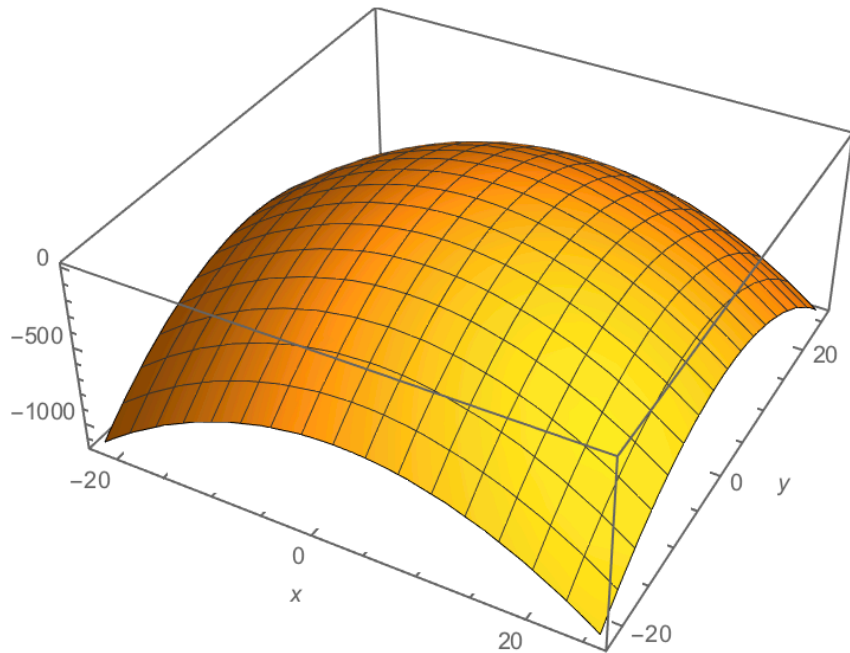
means derivatives are
parallel at maximum.

$$\begin{cases} X = \\ Y = \end{cases}$$

2 equations

Lagrange multipliers example 1

$$f(x, y) = 5 - (x - 2)^2 - (y - 2)^2$$



Contour plot of $f(x, y)$

$$\text{maximize}_{x,y} \quad f(x, y)$$

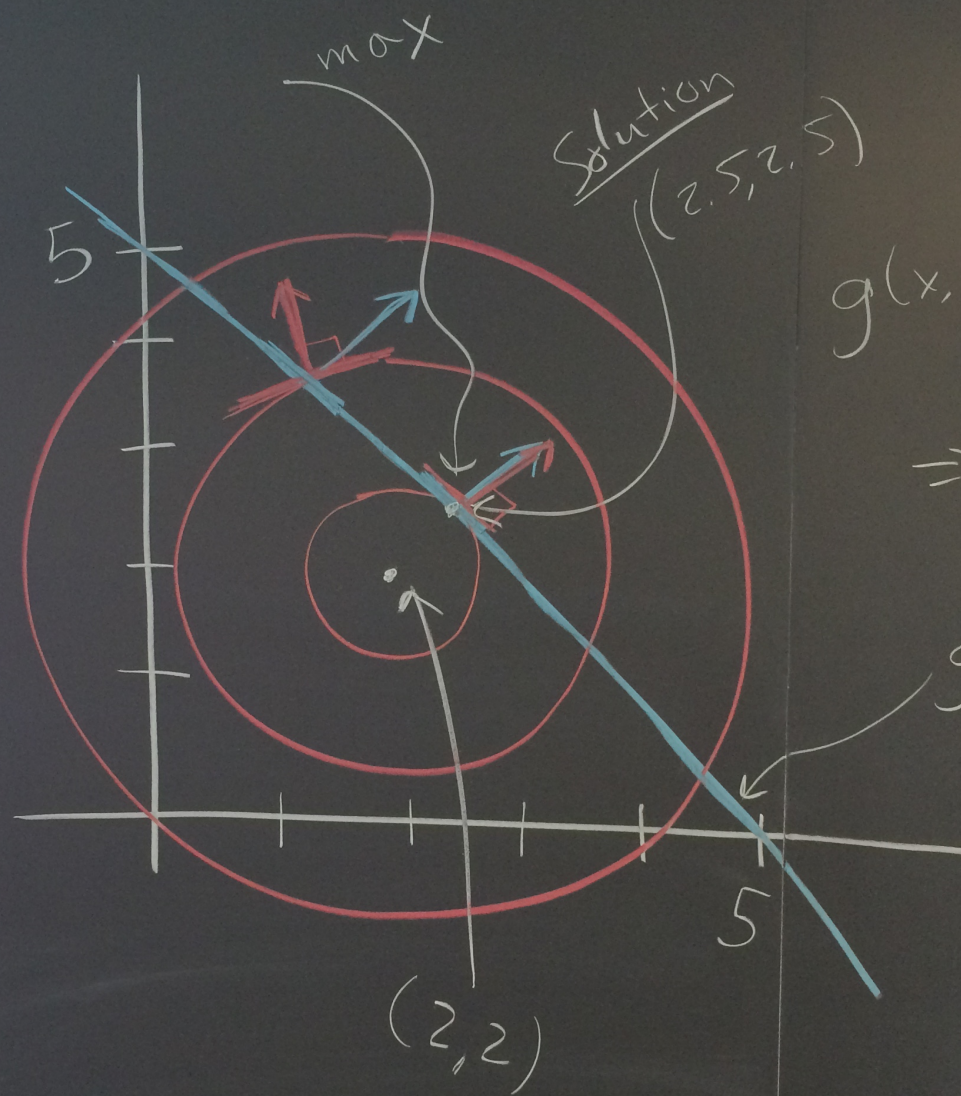
$$\text{s.t.} \quad g(x, y) = 0$$

$$g(x, y) = -5 + x + y$$

SOLUTION:
Normals (and
derivatives)
parallel
 $x = 2.5$
 $y = 2.5$

level curves
of $g(x,y)$

$$g(x,y) = 0$$



Solution
 $(2.5, 2.5)$

$$g(x, y) = -5 + x + y = 0$$

$$\Rightarrow \boxed{x + y = 5}$$

$$g(x, y) = 0$$

$(2, 2)$

Example 1

$$\nabla h = 0 \left\{ \right.$$

$$-5 + x + y = 0$$

$$-2(x-2) = \lambda \cdot 1$$

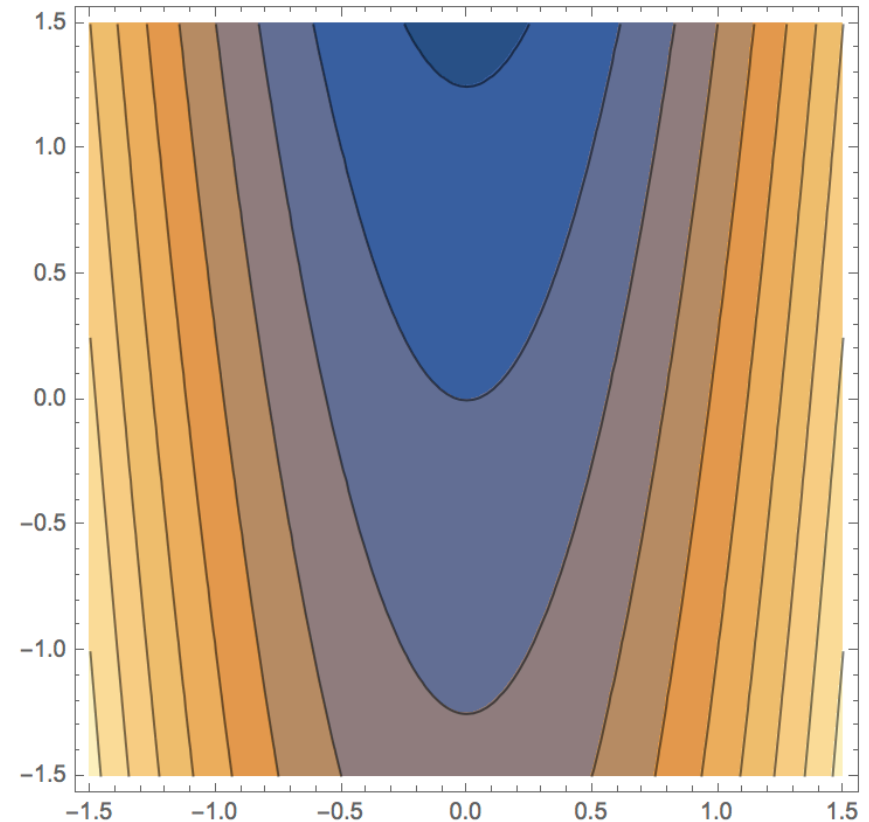
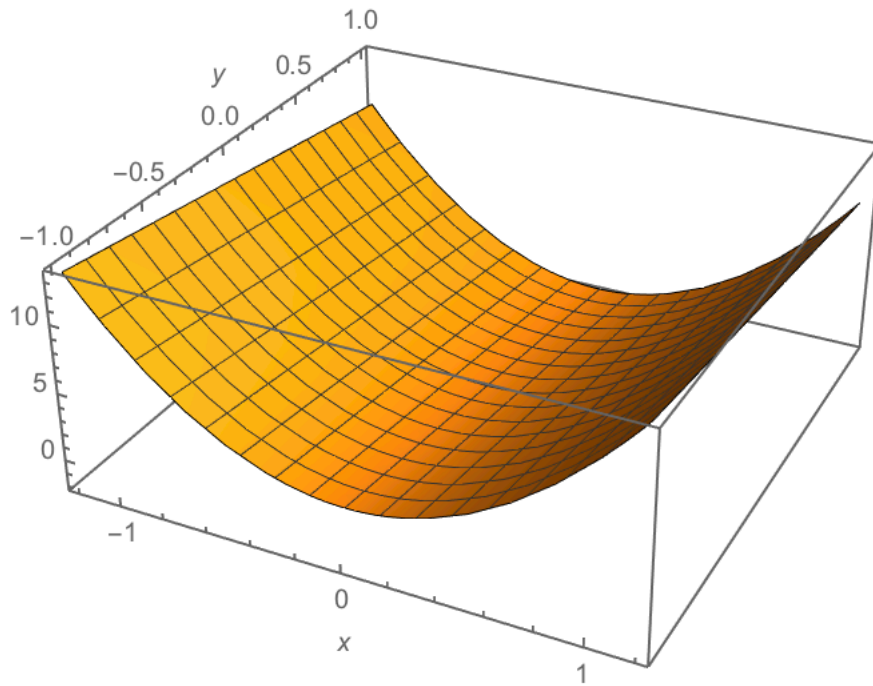
$$-2(y-2) = \lambda \cdot 1$$

$$\begin{aligned} \hookrightarrow x &= -\frac{\lambda}{2} + 2 \\ y &= -\frac{\lambda}{2} + 2 \end{aligned} \left. \vphantom{\begin{aligned} \hookrightarrow x &= -\frac{\lambda}{2} + 2 \\ y &= -\frac{\lambda}{2} + 2 \end{aligned}} \right\} \begin{aligned} &\left(-\frac{\lambda}{2} + 2\right) \cdot 2 = 5 \\ &\Rightarrow \boxed{\lambda = -1} \end{aligned}$$

$$x = 2.5$$

$$y = 2.5$$

Lagrange multipliers example 2

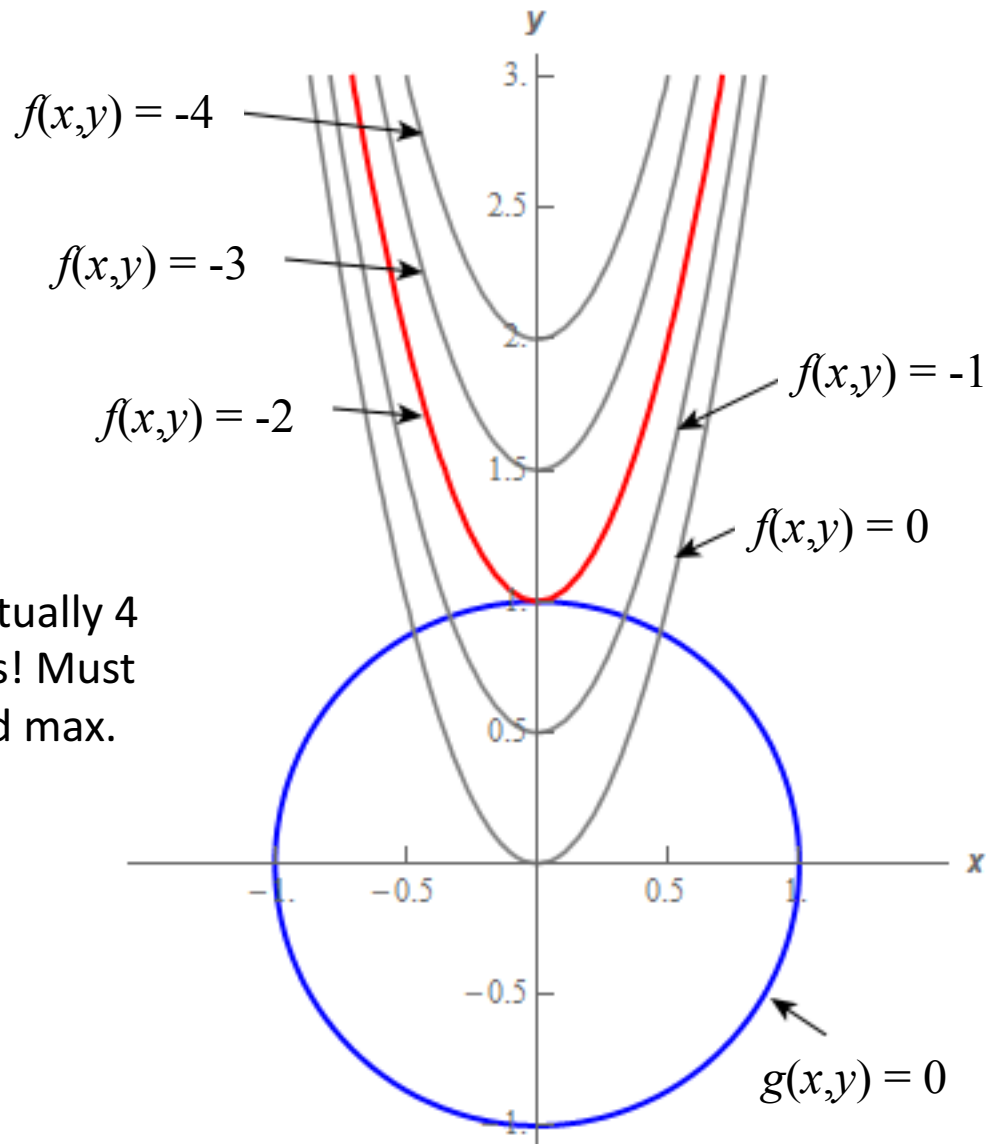


$$\text{maximize}_{x,y} \quad f(x, y)$$

$$s.t. \quad g(x, y) = 0$$

Contour plot of $f(x, y)$

Lagrange multipliers example 2



Note: there are actually 4 potential solutions! Must plug in to f to find max.

Outline for March 25

- Recap SVM introduction and idea of SVMs as an optimization problem
- Detour to Lagrange multipliers as a way to solve optimization problems
- **Reformulation of SVMs using inner products between examples**

Back to
SVMs

$$\min \frac{1}{2} \|\vec{w}\|^2$$

\vec{w}, b

s.t.

$$-y_i(\vec{w} \cdot \vec{x}_i + b) + 1 \leq 0$$

$i = 1, \dots, n$

$$\hat{y} = 1$$

$$L(\vec{w}, b, \vec{\alpha}) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^n \alpha_i [y_i(\vec{w} \cdot \vec{x}_i + b) - 1]$$

α_i is Lagrange multiplier for \vec{x}_i

$\alpha_i = 0$ when \vec{x}_i is not on margin

$\alpha_i > 0$ when \vec{x}_i is on margin
(constraint is active)