CS 360: Machine Learning

Sara Mathieson, Sorelle Friedler

Spring 2024

HAVERFORD COLLEGE

Admin

- **Lab 6** due **Monday March 25**
  - Note: deadline is wrong on github classroom
- Sorelle office hours: **TODAY 3-4pm in H204**
- Sara office hours: Monday 4-5pm in H110

Outline for March 21

- Go over Midterm 1
- Recap perceptron algorithm
- Support Vector Machine theory
- SVM extensions

Outline for March 21

- Go over Midterm 1
- Recap perceptron algorithm
- Support Vector Machine theory
- SVM extensions

(not posted online)

Outline for March 21

- Go over Midterm 1
- Recap perceptron algorithm
- Support Vector Machine theory
- SVM extensions

Perception partner exercise

1. What is the goal of the perceptron algorithm? Circle all that apply:

   (a) predict a continuous outcome

   (b) quantify how important each feature is for predicting the outcome

   (c) create a linear decision boundary between positives and negatives

   (d) obtain the probability of a positive label for each test example

2. *True or False*: The perceptron algorithm was inspired by how neurons are activated in our brains.

3. Say at some point in the perceptron algorithm I have $\vec{w} = [3, -1, 2]^T$ and $\vec{x} = [1, 2, -2]^T$. What label would we predict for $\vec{x}$?

4. In the example above, say the true label is -1. How would the weights be updated when using this point?

Perception partner exercise

1. What is the goal of the perceptron algorithm? Circle all that apply:

   (a) predict a continuous outcome

   (b) quantify how important each feature is for predicting the outcome

   (c) create a linear decision boundary between positives and negatives

   (d) obtain the probability of a positive label for each test example

Perception partner exercise

1. What is the goal of the perceptron algorithm? Circle all that apply:

   (a) predict a continuous outcome

   (b) quantify how important each feature is for predicting the outcome

   (c) create a linear decision boundary between positives and negatives

   (d) obtain the probability of a positive label for each test example

2. *True or False*: The perceptron algorithm was inspired by how neurons are activated in our brains.

Perception partner exercise

1. What is the goal of the perceptron algorithm? Circle all that apply:

   (a) predict a continuous outcome

   (b) quantify how important each feature is for predicting the outcome

   (c) create a linear decision boundary between positives and negatives

   (d) obtain the probability of a positive label for each test example

2. *True or False*: The perceptron algorithm was inspired by how neurons are activated in our brains.

   True

3. Say at some point in the perceptron algorithm I have $\vec{w} = [3, -1, 2]^T$ and $\vec{x} = [1, 2, -2]^T$. What label would we predict for $\vec{x}$?

Perception partner exercise

1. What is the goal of the perceptron algorithm? Circle all that apply:

   (a) predict a continuous outcome

   $\boxed{\text{(b) quantify how important each feature is for predicting the outcome}}$

   $\boxed{\text{(c) create a linear decision boundary between positives and negatives}}$

   (d) obtain the probability of a positive label for each test example

2. *True or False*: The perceptron algorithm was inspired by how neurons are activated in our brains.

   True

3. Say at some point in the perceptron algorithm I have $\vec{w} = [3, -1, 2]^T$ and $\vec{x} = [1, 2, -2]^T$. What label would we predict for $\vec{x}$?

   Dot product = -3 => predict label -1

4. In the example above, say the true label is -1. How would the weights be updated when using this point?

Perception partner exercise

1. What is the goal of the perceptron algorithm? Circle all that apply:

   (a) predict a continuous outcome

   (b) quantify how important each feature is for predicting the outcome

   (c) create a linear decision boundary between positives and negatives

   (d) obtain the probability of a positive label for each test example

2. *True or False*: The perceptron algorithm was inspired by how neurons are activated in our brains.

   True

3. Say at some point in the perceptron algorithm I have $\vec{w} = [3, -1, 2]^T$ and $\vec{x} = [1, 2, -2]^T$. What label would we predict for $\vec{x}$?

   Dot product = -3 => predict label -1

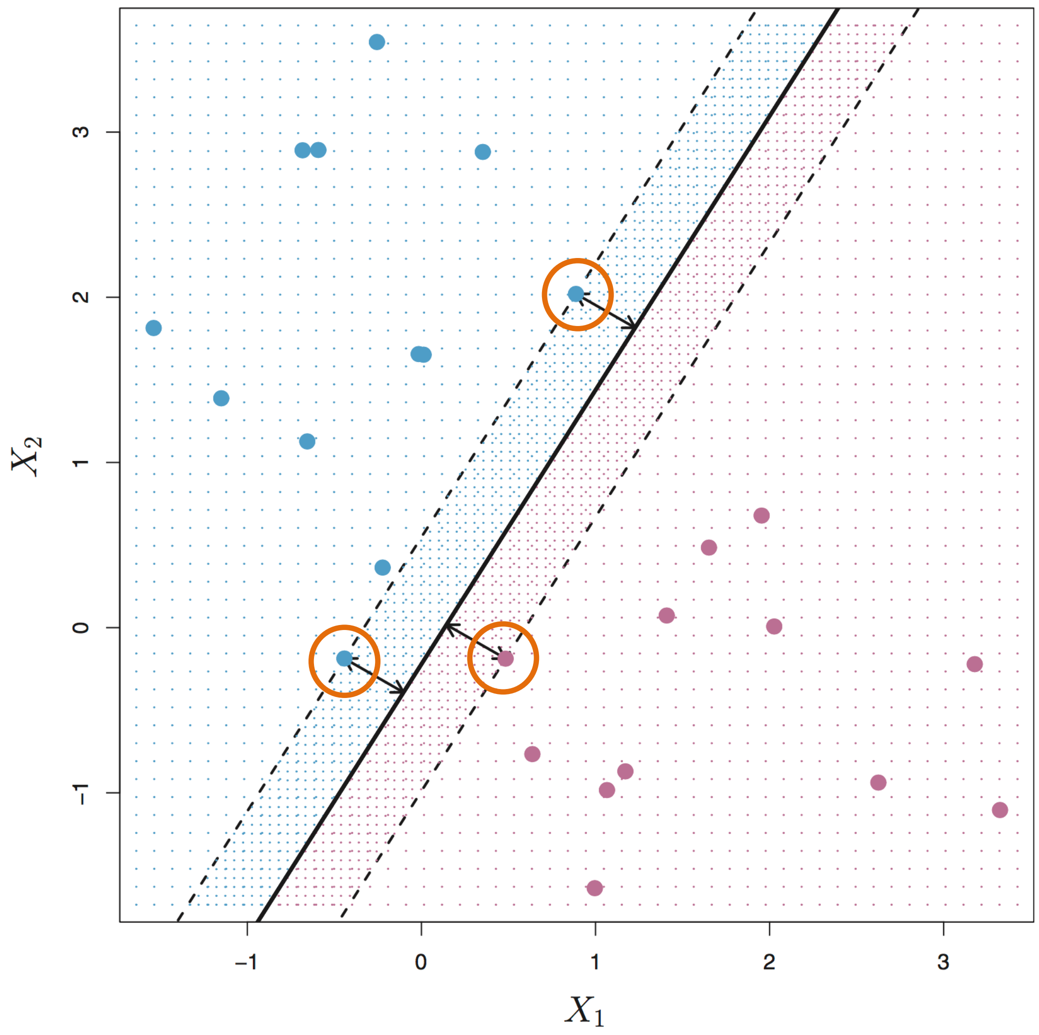
4. In the example above, say the true label is -1. How would the weights be updated when using this point?

   No weight update!

Outline for March 21

- Go over Midterm 1
- Recap perceptron algorithm
- Support Vector Machine theory
- SVM extensions

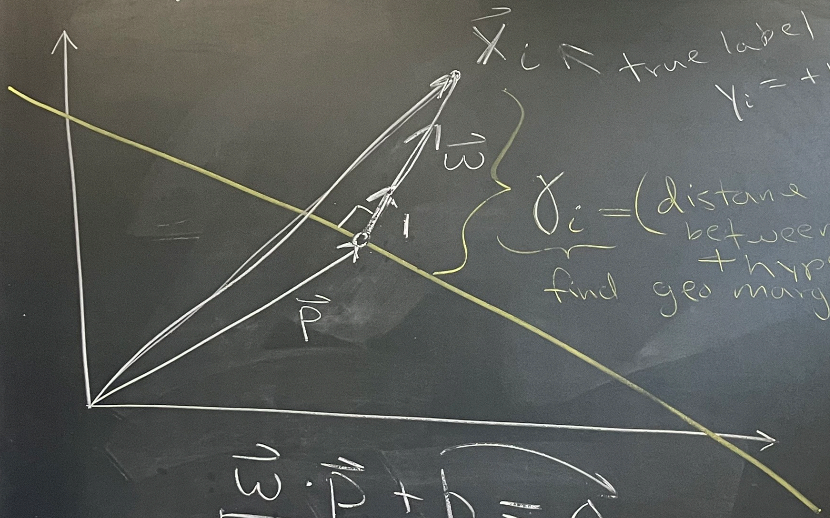
Datapoints that lie on the margin are called "support vectors"



Support vectors

Modified from ISL: Figure 9.3

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

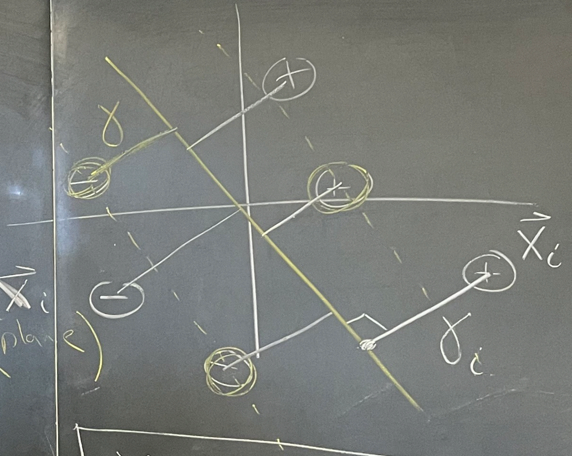


$\gamma_i = $ (distance between $\vec{x}_i$ + hyperplane)

find geo margin

$\vec{w}\cdot\vec{p} + b = 0$

$\vec{p}$ is on hyperplane



$\boxed{\gamma = \min_{i=1\ldots n} \gamma_i}$

↑ overall geometric margin

↑ geometric margin for example i

goal

maximize γ

$\|\vec{w}\| = \sqrt{w_1^2 + w_2^2 + \cdots w_p^2}$

norm

magnitude

$$\vec{w}\cdot\left(\vec{p} + \gamma_i \underbrace{\frac{\vec{w}}{\|\vec{w}\|}}_{\text{unit vector}} = \vec{x}_i\right) \Big\} \text{ solve for } \gamma_i$$

y_i

$$\overset{-b}{\cancel{\vec{w}\cdot\vec{p}}} + \gamma_i y_i \frac{\overbrace{\vec{w}\cdot\vec{w}}^{\|\vec{w}\|^{\cancel{2}}}}{\cancel{\|\vec{w}\|}} = \vec{w}\cdot\vec{x}_i$$

$$-b + \gamma_i y_i \|\vec{w}\| = \vec{w}\cdot\vec{x}_i$$

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

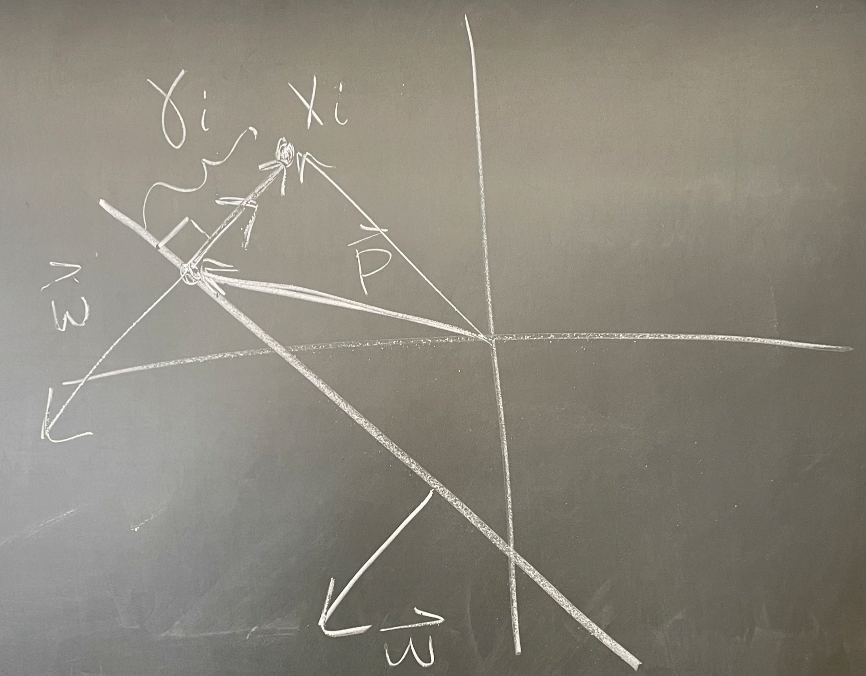
geometric margin

$$y_i = \frac{1}{y_i}$$

functional margin

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

make bigger by making $\|\vec{w}\|$ bigger



γi
xi
p
w
w

Functional and Geometric Margins

SVM classifier:
(same as perceptron)

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

Functional and Geometric Margins

SVM classifier:
(same as perceptron)

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

Functional Margin:

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

Functional and Geometric Margins

SVM classifier:
(same as perceptron)

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

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)

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

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:

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

Optimization Problem: try 1

Goal: maximize the minimum distance between example and hyperplane

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

Optimization Problem: try 1

Goal: maximize the minimum distance between example and hyperplane

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

Formulation: optimize a function with respect to a constraint

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

(force functional and geometric margin to be equal)

Optimization Problem: try 2

Idea: substitute functional margin
divided by magnitude of weight vector

$$\max_{\hat{\gamma},\vec{w},b} \quad \frac{\hat{\gamma}}{||\vec{w}||}$$

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

(gets rid of non-convex constraint)

Optimization Problem: try 3

Idea: put arbitrary constraint on functional margin

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

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

Optimization Problem: try 3

Idea: put arbitrary constraint on functional margin

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

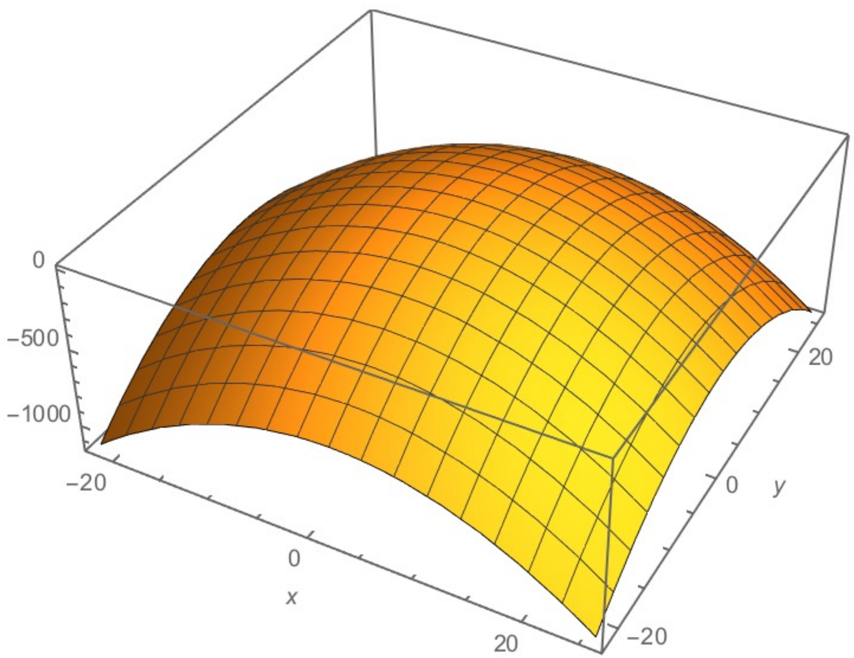
$$\min_{\vec{w},b} \quad \frac{1}{2}\|\vec{w}\|^2$$

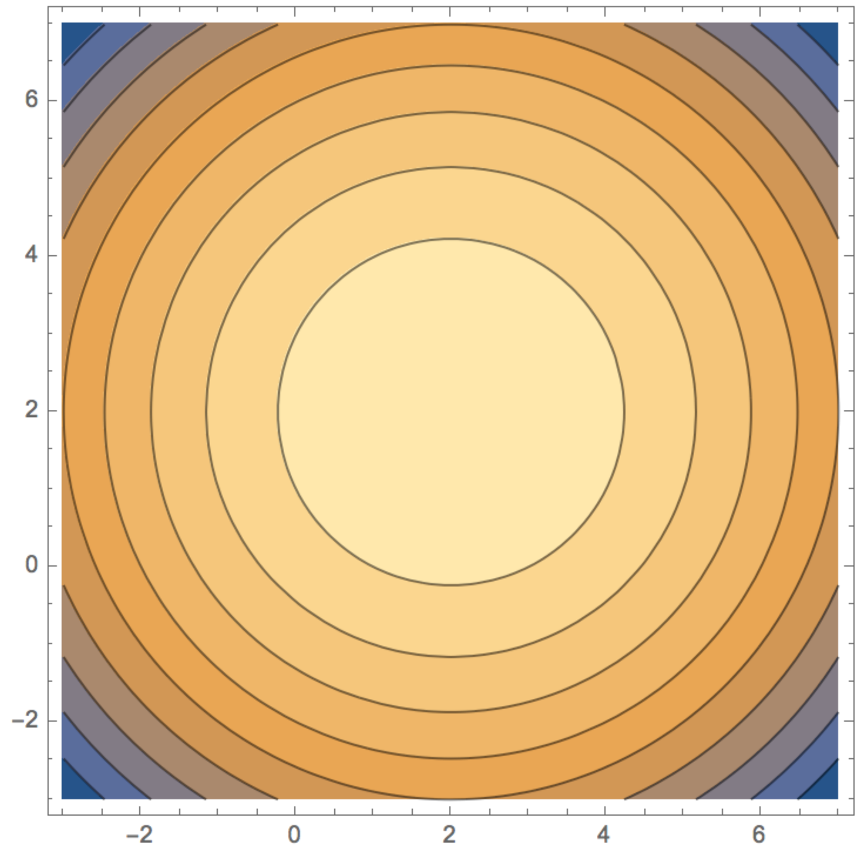
$$\text{s.t.} \quad y_i(\vec{w}\cdot\vec{x}_i + b) \geq 1, \quad i = 1, \cdots, n$$

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

Lagrange multipliers example

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





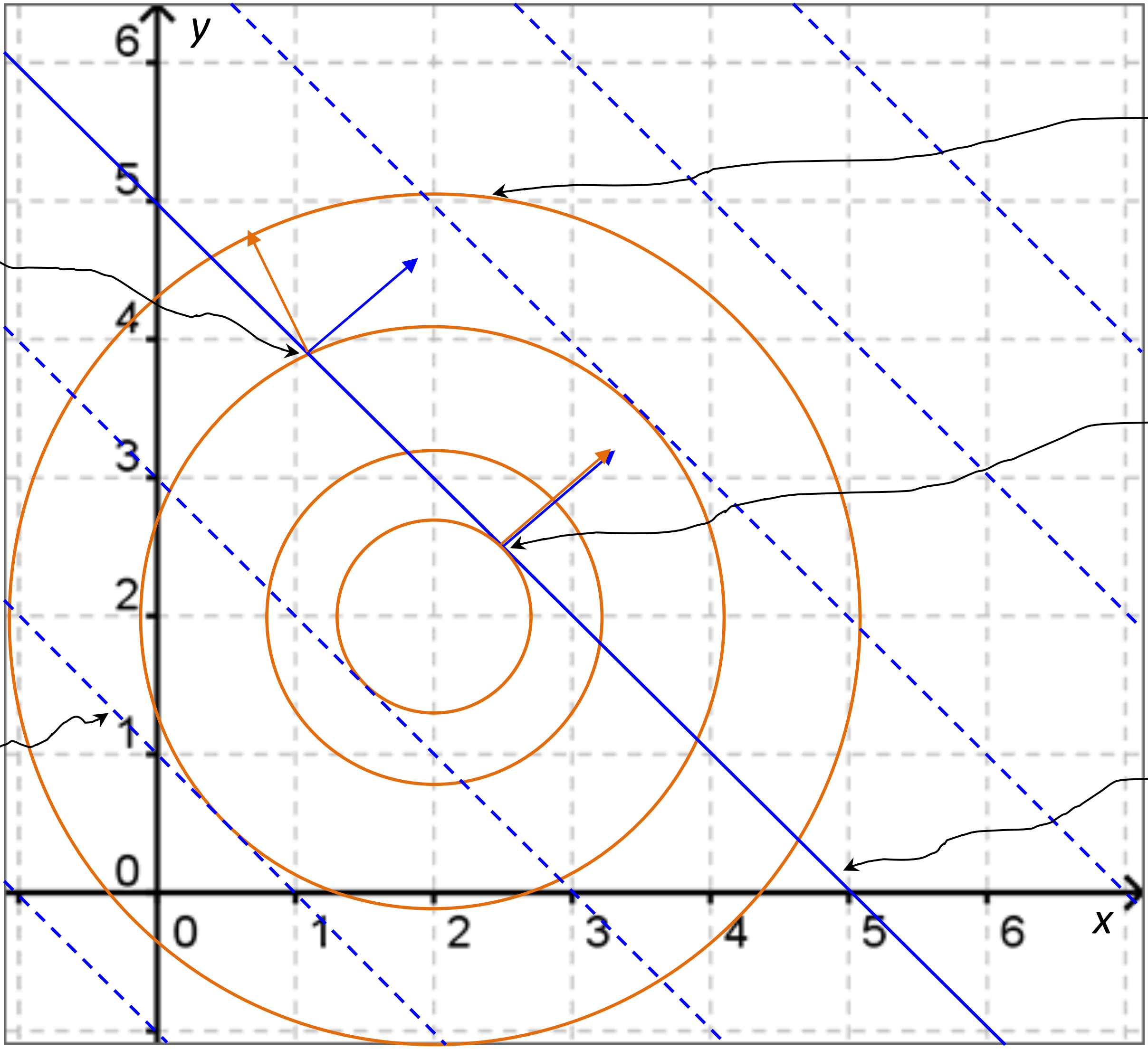
Contour plot of *f*(*x*,*y*)

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

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

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

Lagrange multipliers example



level curves of $f(x,y)$

Normals (and derivatives) **not** parallel

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

level curves of $g(x,y)$

$g(x,y) = 0$

Outline for March 21

- Go over Midterm 1
- Recap perceptron algorithm
- Support Vector Machine theory
- SVM extensions