

CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019



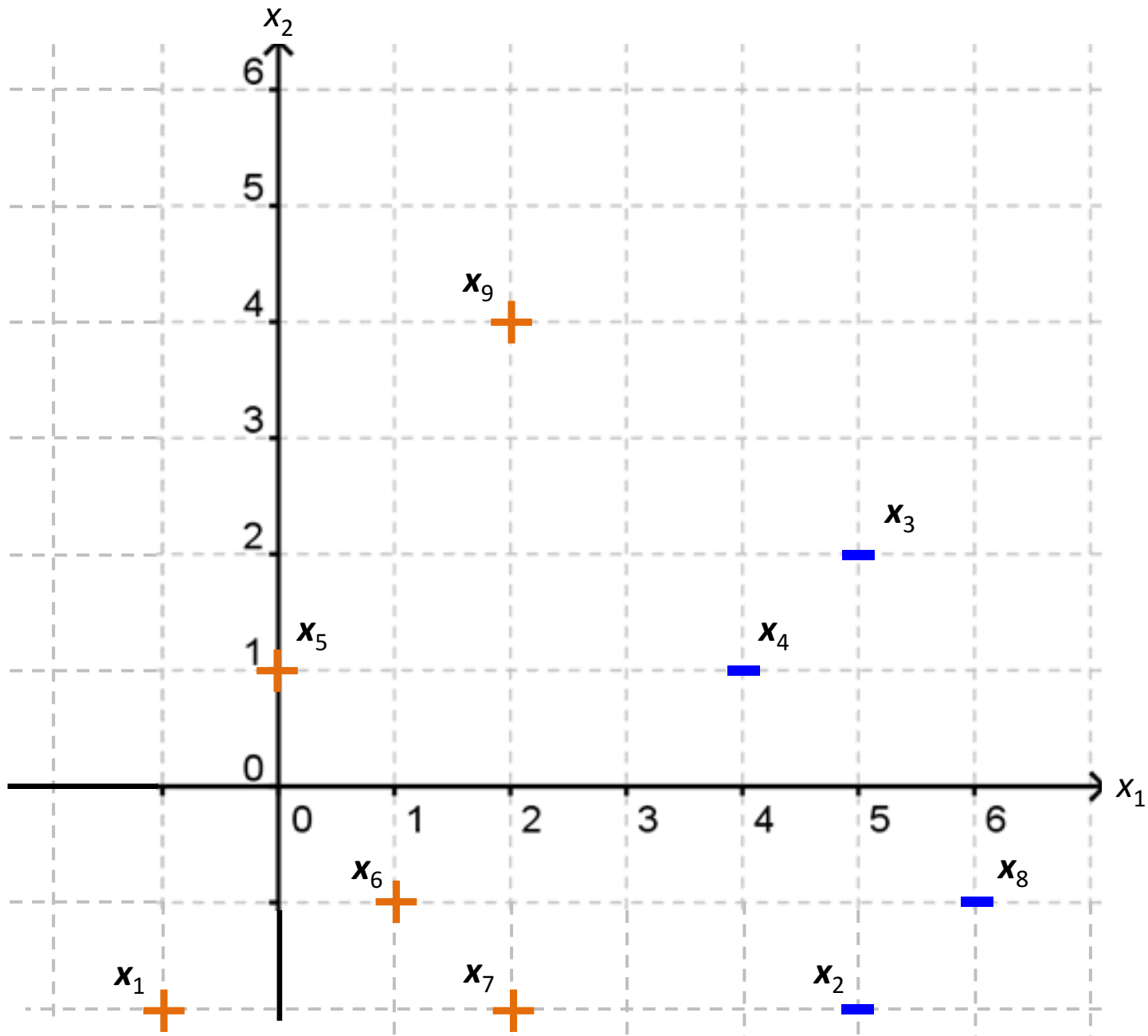
Outline for April 1

- Handout 12 followup
 - Relating Lagrangian and geometry
 - Begin: Neural Networks
 - Introduction
 - Notation and diagrams
 - Backpropagation
- Lab 6 due Friday
 - Lab 7 released Friday
 - Lab 6 check-in Wednesday
 - Office hours today 12:30-2pm

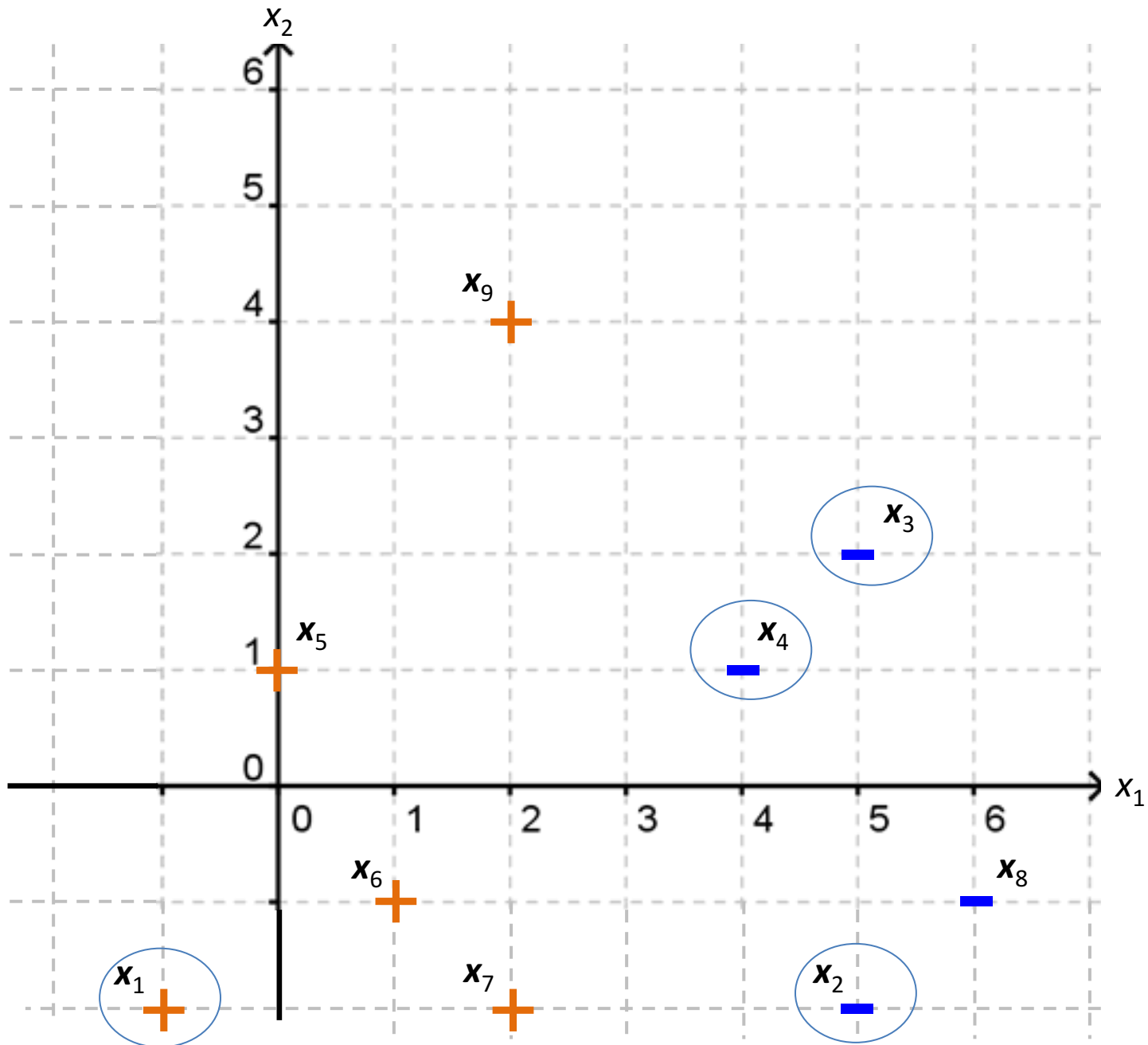
Outline for April 1

- Handout 12 followup
- Relating Lagrangian and geometry
- Begin: Neural Networks
 - Introduction
 - Notation and diagrams
 - Backpropagation

Handout 12 Example



Meta-optimization: example



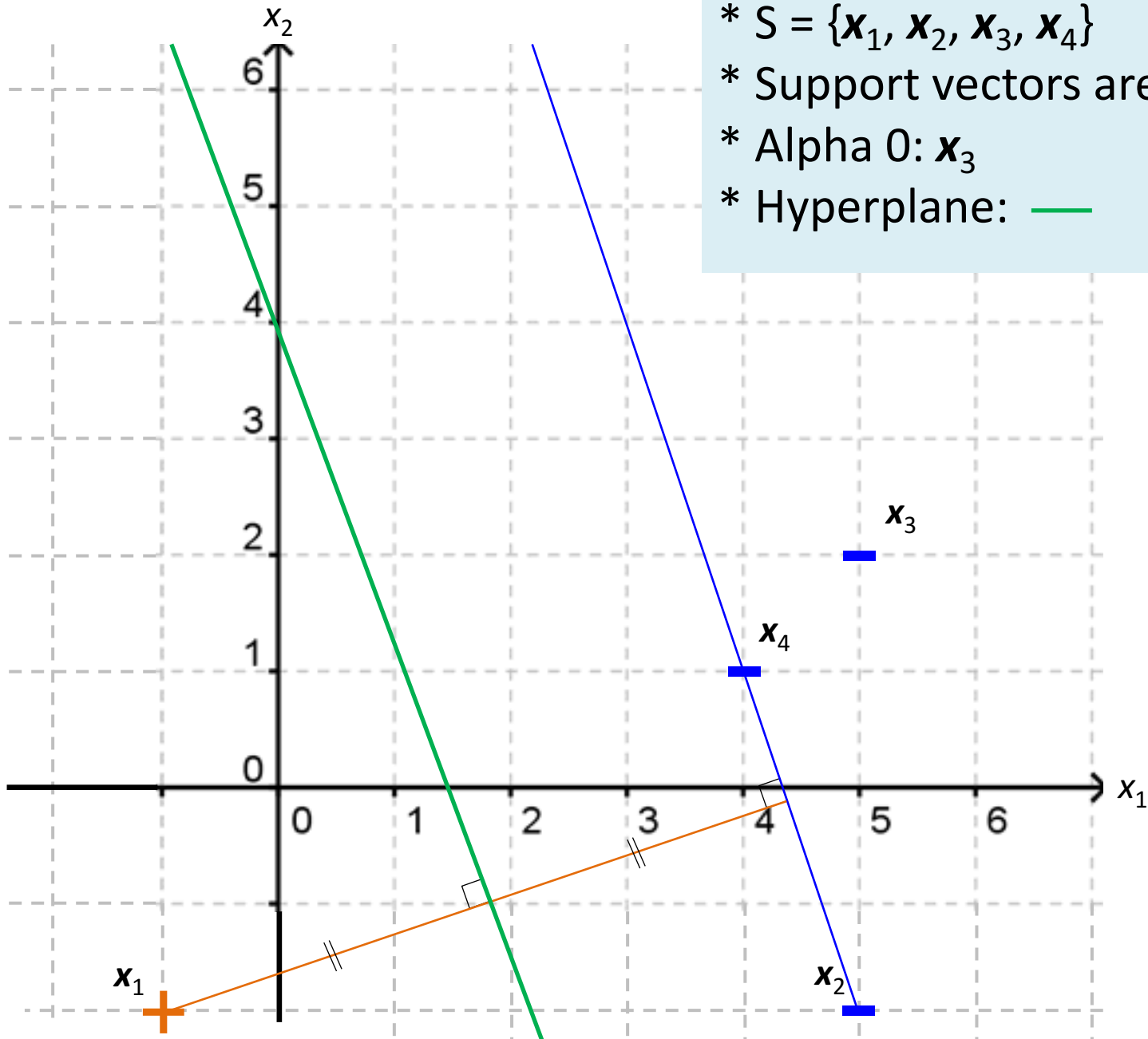
Round 1:

* $S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$

* Support vectors are: $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4$

* Alpha 0: \mathbf{x}_3

* Hyperplane: —



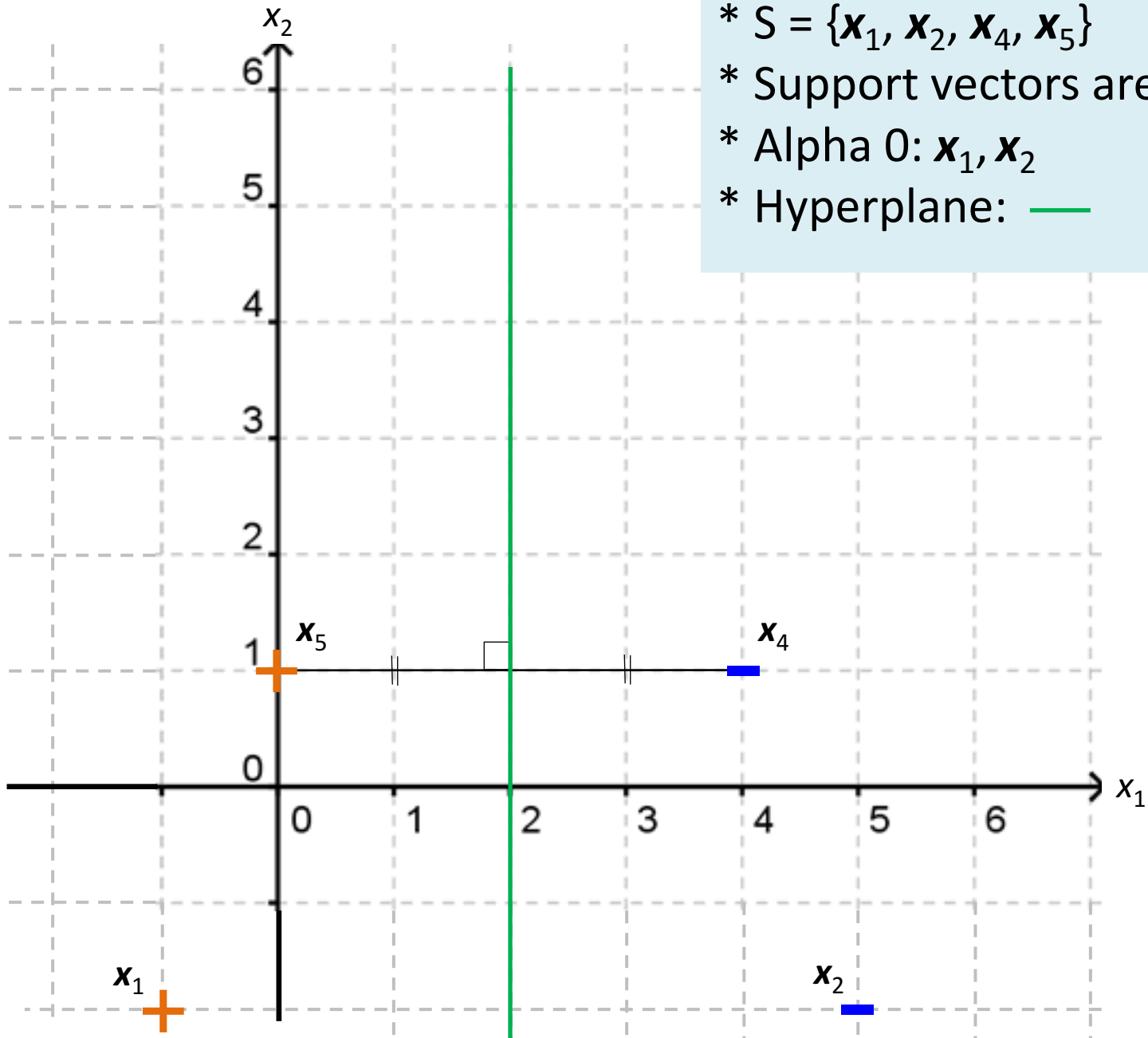
Round 1:

* $S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4, \mathbf{x}_5\}$

* Support vectors are: $\mathbf{x}_4, \mathbf{x}_5$

* Alpha 0: $\mathbf{x}_1, \mathbf{x}_2$

* Hyperplane: —



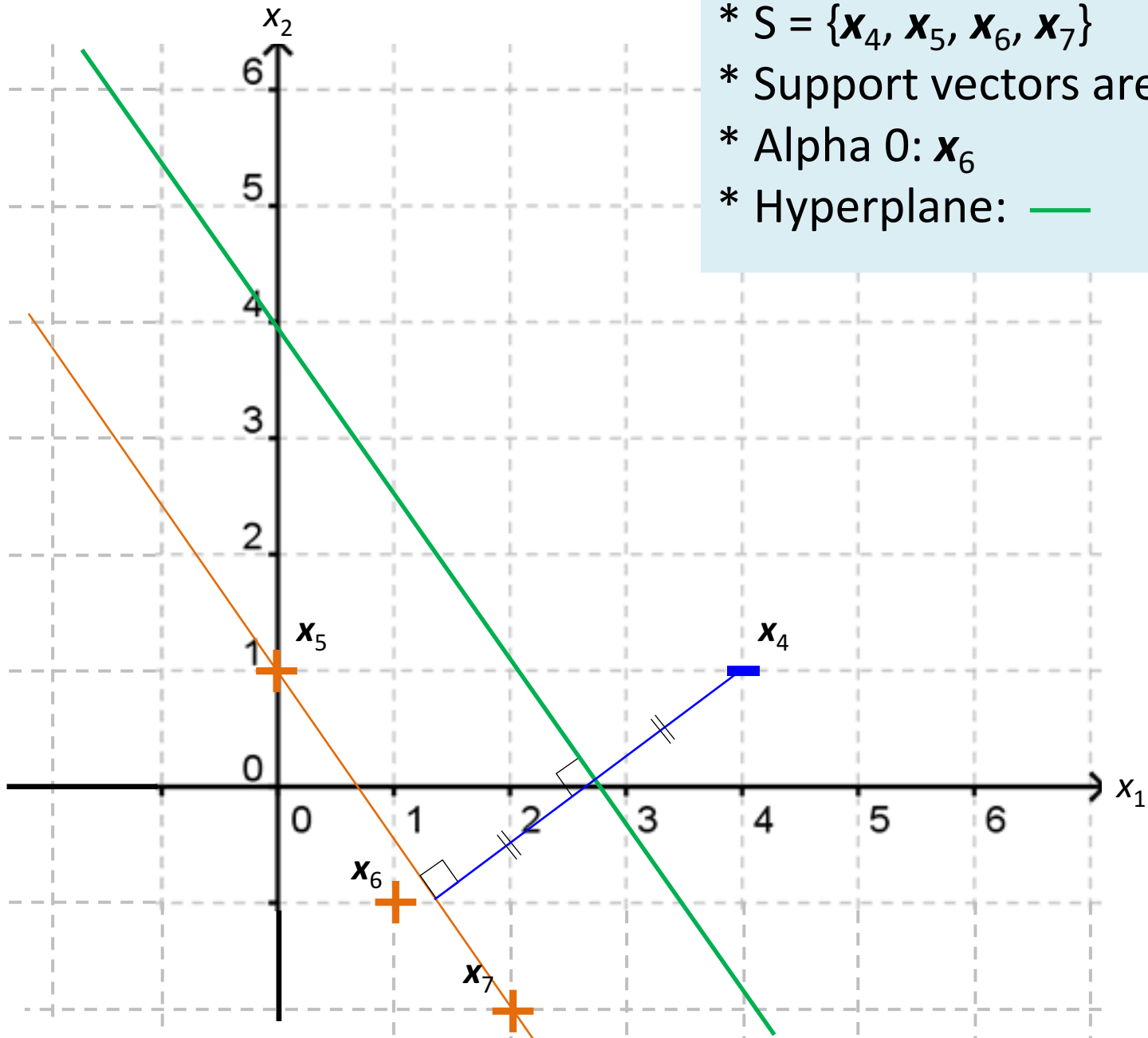
Round 3:

* $S = \{\mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7\}$

* Support vectors are: $\mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_7$

* Alpha 0: \mathbf{x}_6

* Hyperplane: —



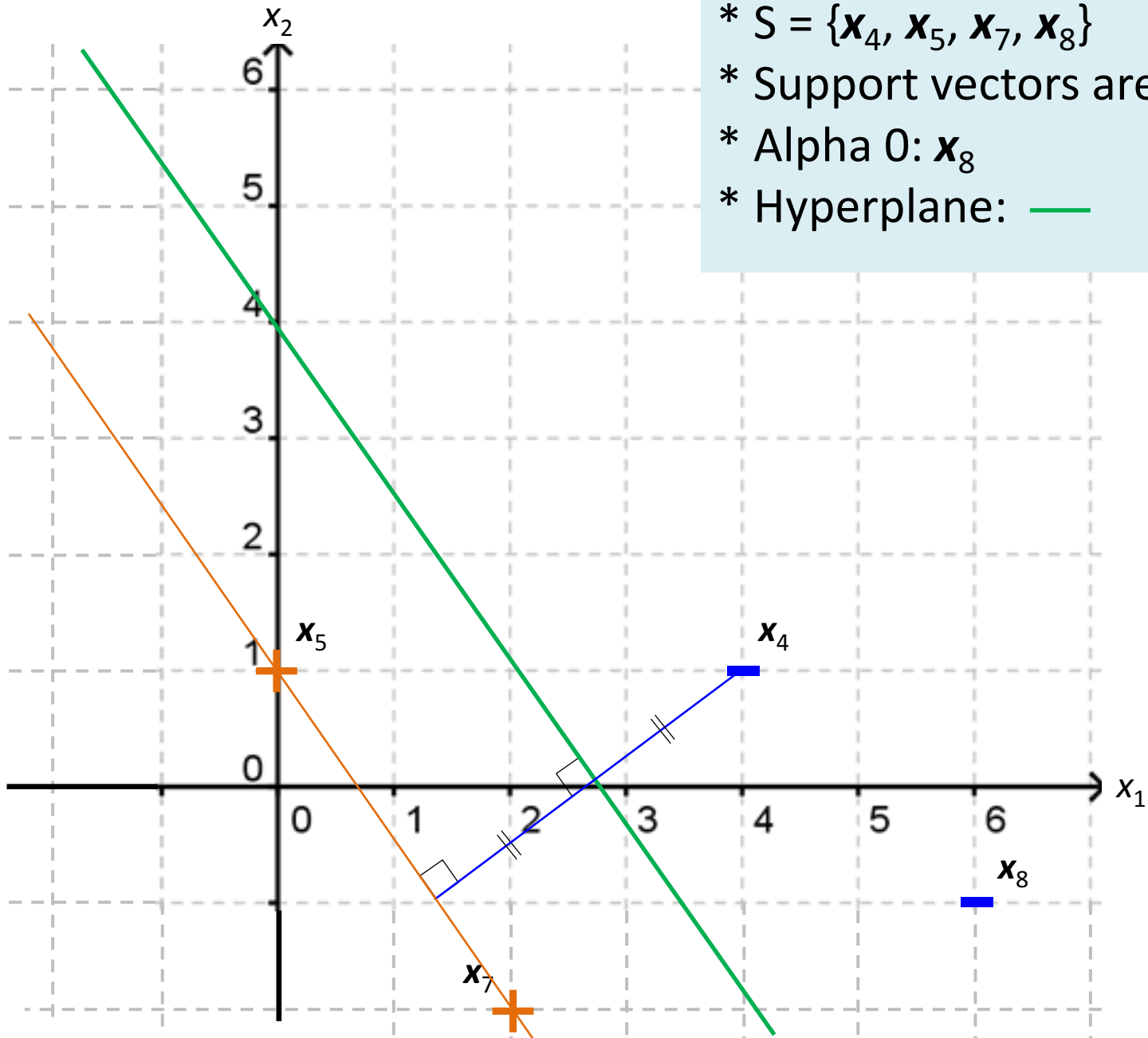
Round 4:

* $S = \{\mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_7, \mathbf{x}_8\}$

* Support vectors are: $\mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_7$

* Alpha 0: \mathbf{x}_8

* Hyperplane: —



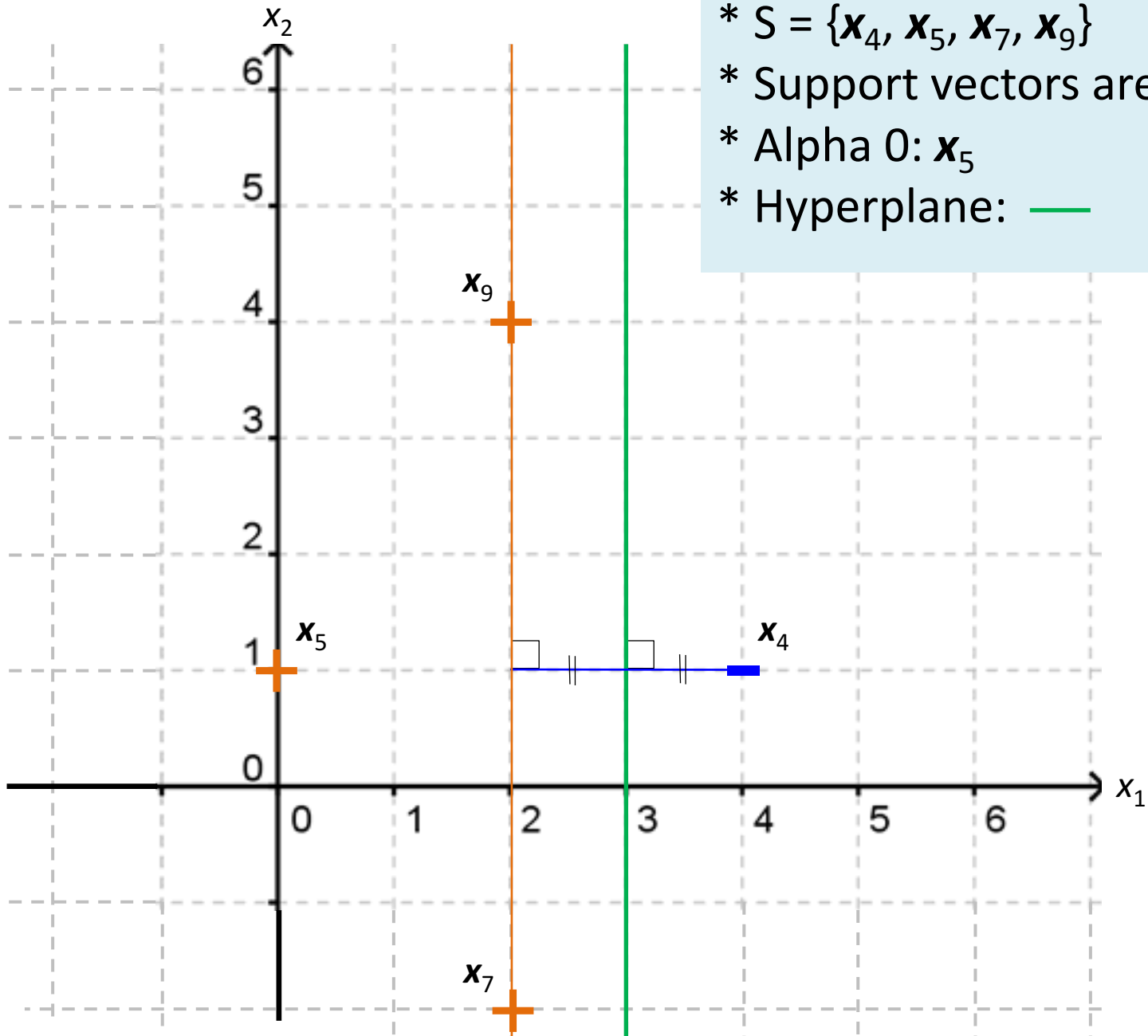
Round 5:

* $S = \{\mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_7, \mathbf{x}_9\}$

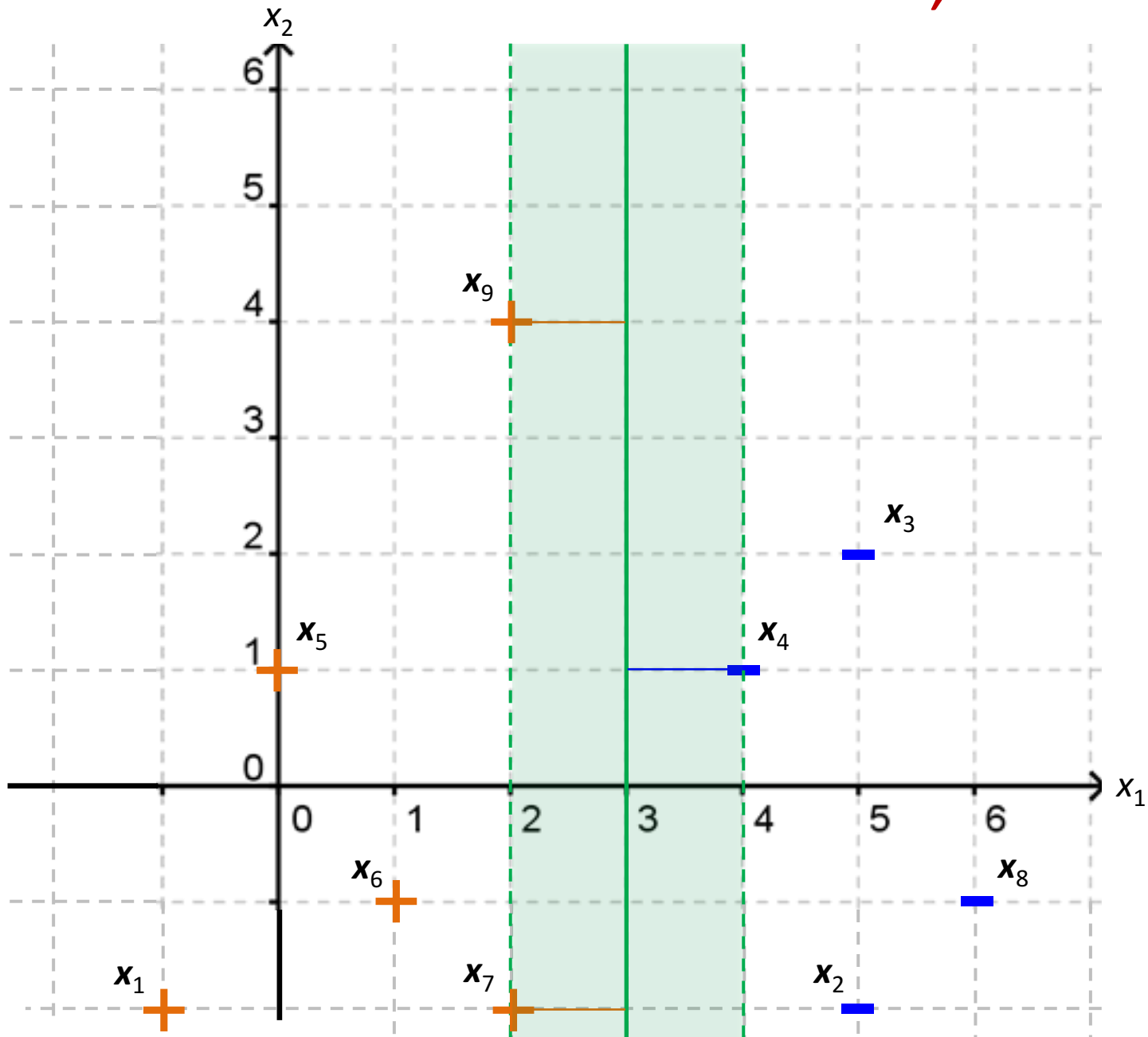
* Support vectors are: $\mathbf{x}_4, \mathbf{x}_7, \mathbf{x}_9$

* Alpha 0: \mathbf{x}_5

* Hyperplane: —

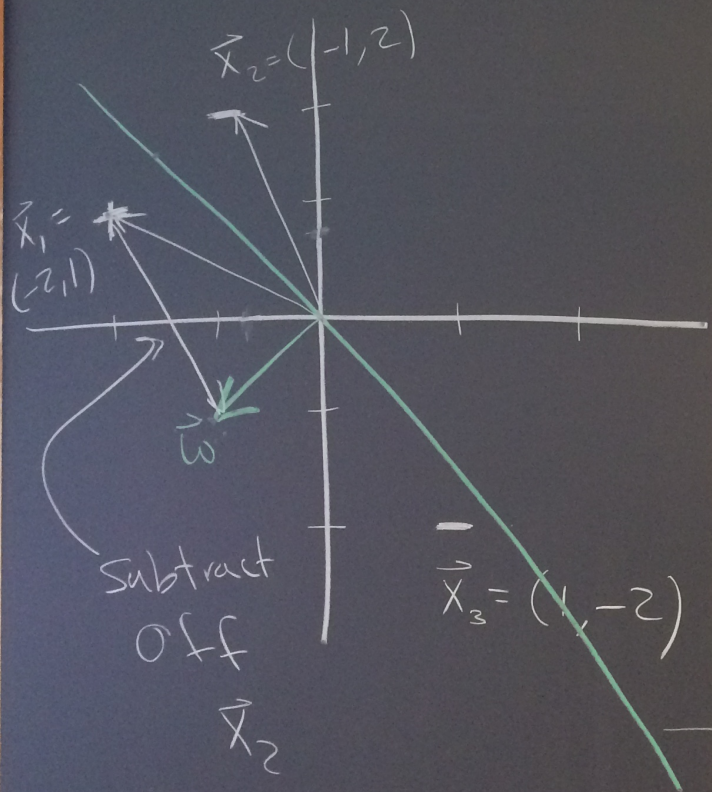


Handout 12, Final Solution



Outline for April 1

- Handout 12 followup
- **Relating Lagrangian and geometry**
- Begin: Neural Networks
 - Introduction
 - Notation and diagrams
 - Backpropagation



$$\vec{x}_1 \cdot \vec{x}_1 = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 1 \end{bmatrix} = 5 \quad y_1 y_1 = 1$$

$$\vec{x}_2 \cdot \vec{x}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} = 5 \quad y_2 y_2 = 1$$

$$\vec{x}_3 \cdot \vec{x}_3 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -2 \end{bmatrix} = 5 \quad y_3 y_3 = 1$$

$$\vec{x}_1 \cdot \vec{x}_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} = 4 \quad y_1 y_2 = -1$$

$$\vec{x}_1 \cdot \vec{x}_3 = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -2 \end{bmatrix} = -4 \quad y_1 y_3 = -1$$

$$\vec{x}_2 \cdot \vec{x}_3 = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -2 \end{bmatrix} = -5 \quad y_2 y_3 = 1$$

$$\max_{\vec{\alpha}} W(\vec{\alpha}) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \underbrace{\alpha_i \alpha_j y_i y_j}_{\text{positive} \Rightarrow \text{some } \alpha \text{ is } 0} \underbrace{\vec{x}_i \cdot \vec{x}_j}_{\text{negative} \Rightarrow \text{need to change some } \alpha}$$

$$\boxed{\alpha_3 = 0}$$

$$\boxed{\alpha_i \geq 0}$$

positive \Rightarrow some α is 0
negative \Rightarrow need to change some α

$$\boxed{\sum \alpha_i y_i = 0}$$

$$\Rightarrow \boxed{\alpha_1 = \alpha_2 = \alpha}$$

$$W(\alpha_1, \alpha_2) = \alpha_1 + \alpha_2 - \frac{1}{2} (5\alpha_1^2 + 5\alpha_2^2 - 4\alpha_1\alpha_2 \cdot 2)$$

$$W(\alpha) = 2\alpha - \frac{1}{2} (10\alpha^2 - 8\alpha^2)$$

$$\boxed{W(\alpha) = 2\alpha - \alpha^2}$$

exercise: $\alpha_1 = 1$, solve for α_2 & α_3

Derivative!

$$W'(\alpha) = 2 - 2\alpha$$

$$\boxed{\alpha^* = 1}$$

$$\sum \alpha_i y_i = 0$$

$$\Rightarrow \boxed{\alpha_1 = \alpha_2} = \alpha$$

Derivative!

$$W'(\alpha) = 2 - 2\alpha = 0$$

$$\boxed{\alpha^* = 1}$$

$$\vec{w} = \sum \alpha_i y_i \vec{x}_i$$

$$\vec{w} = \begin{bmatrix} -2 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

$$\vec{w} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

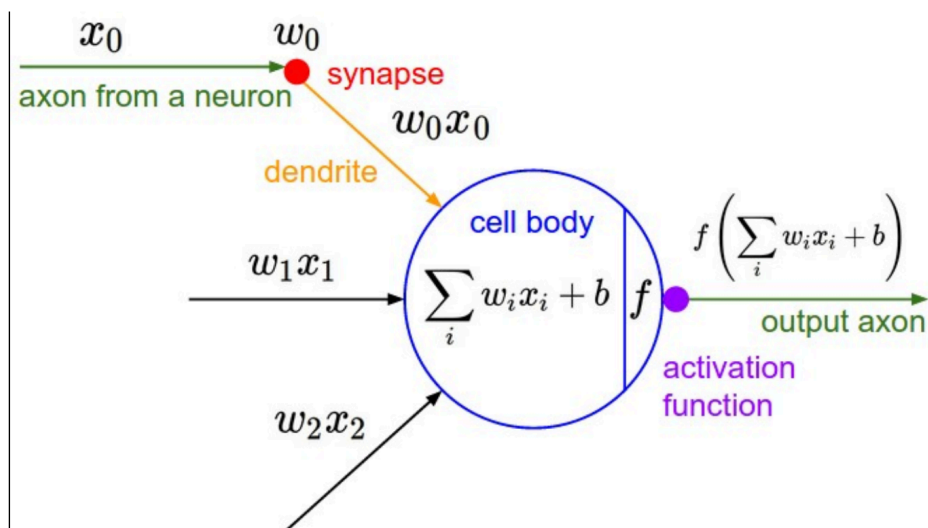
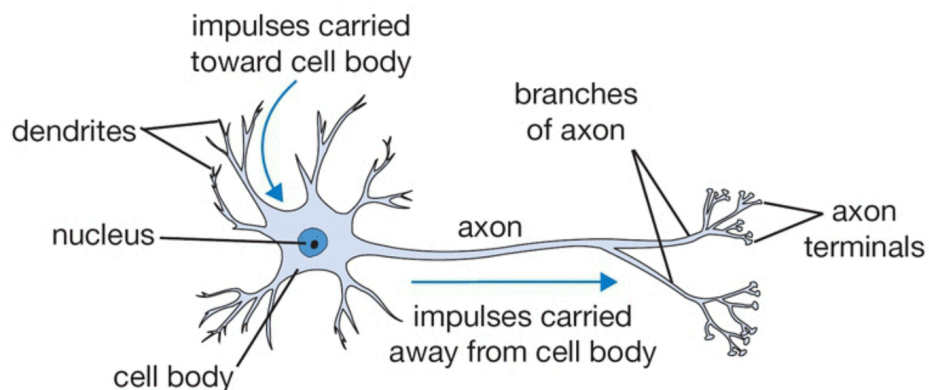
Outline for April 1

- Handout 12 followup
- Relating Lagrangian and geometry
- **Begin: Neural Networks**
 - Introduction
 - Notation and diagrams
 - Backpropagation

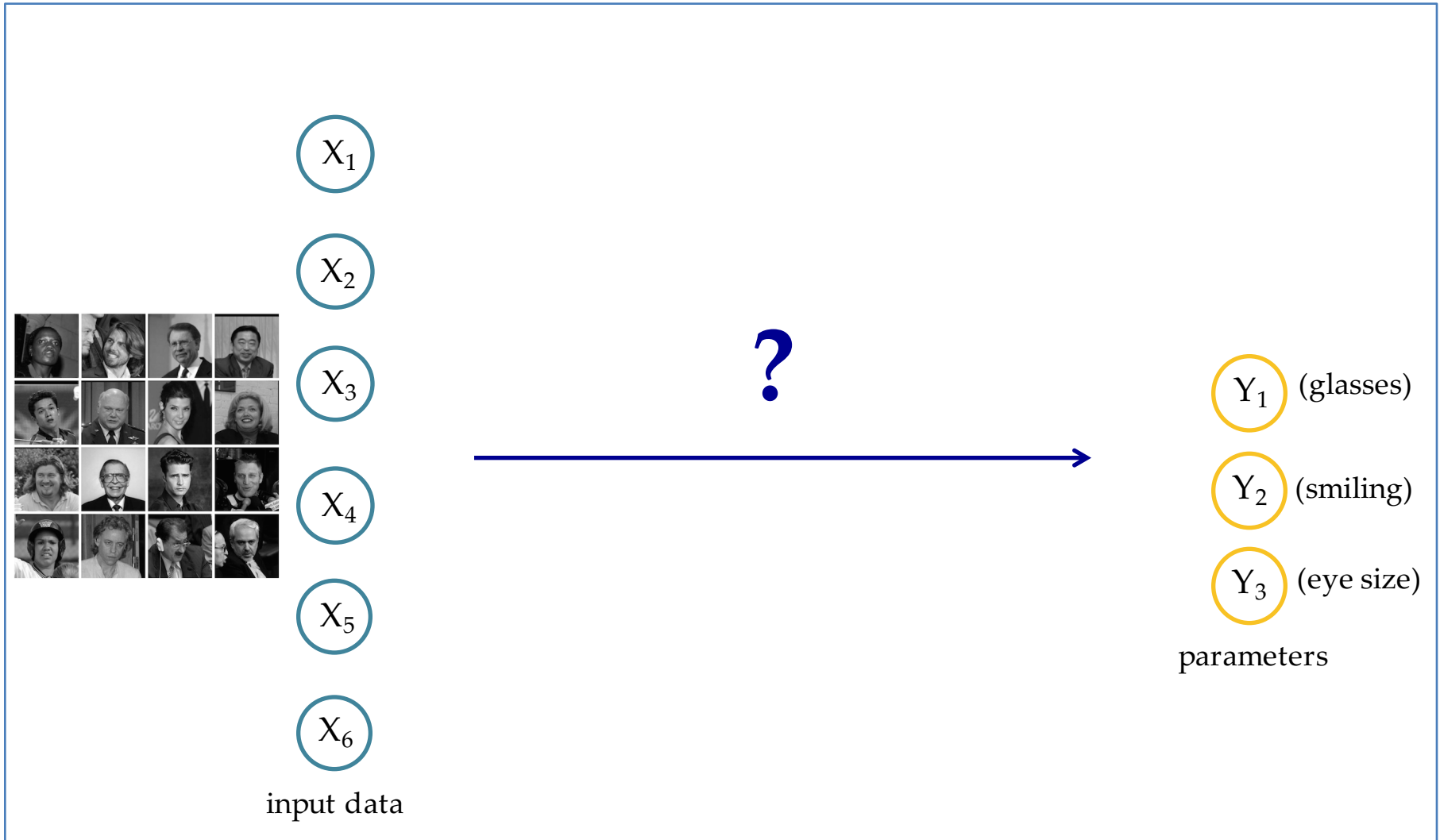
Disadvantages of SVMs

- Difficult to choose a kernel function
- Does not naturally take into account the correlations between features
- Hard to understand and interpret what the model has learned

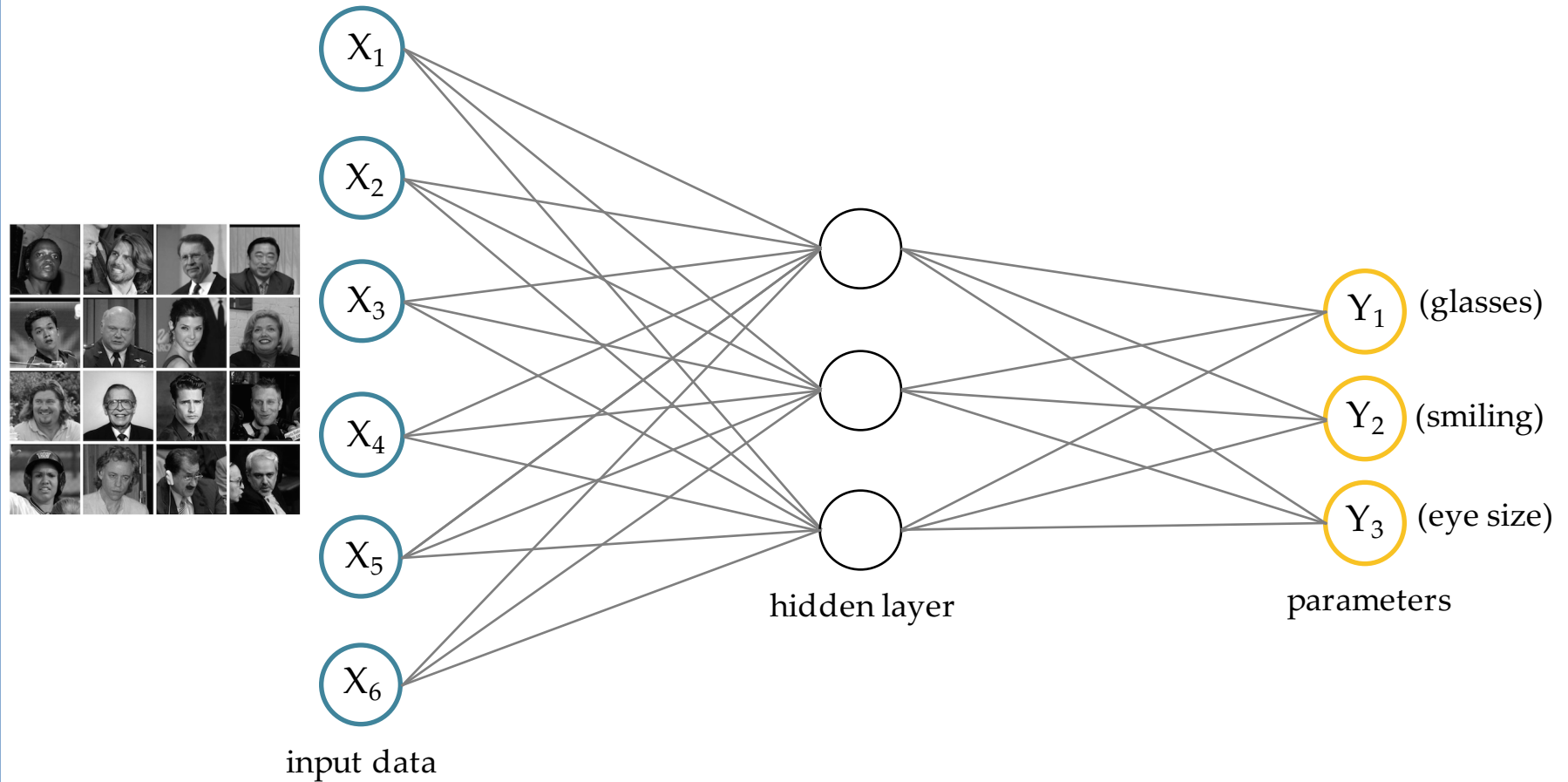
Biological Inspiration



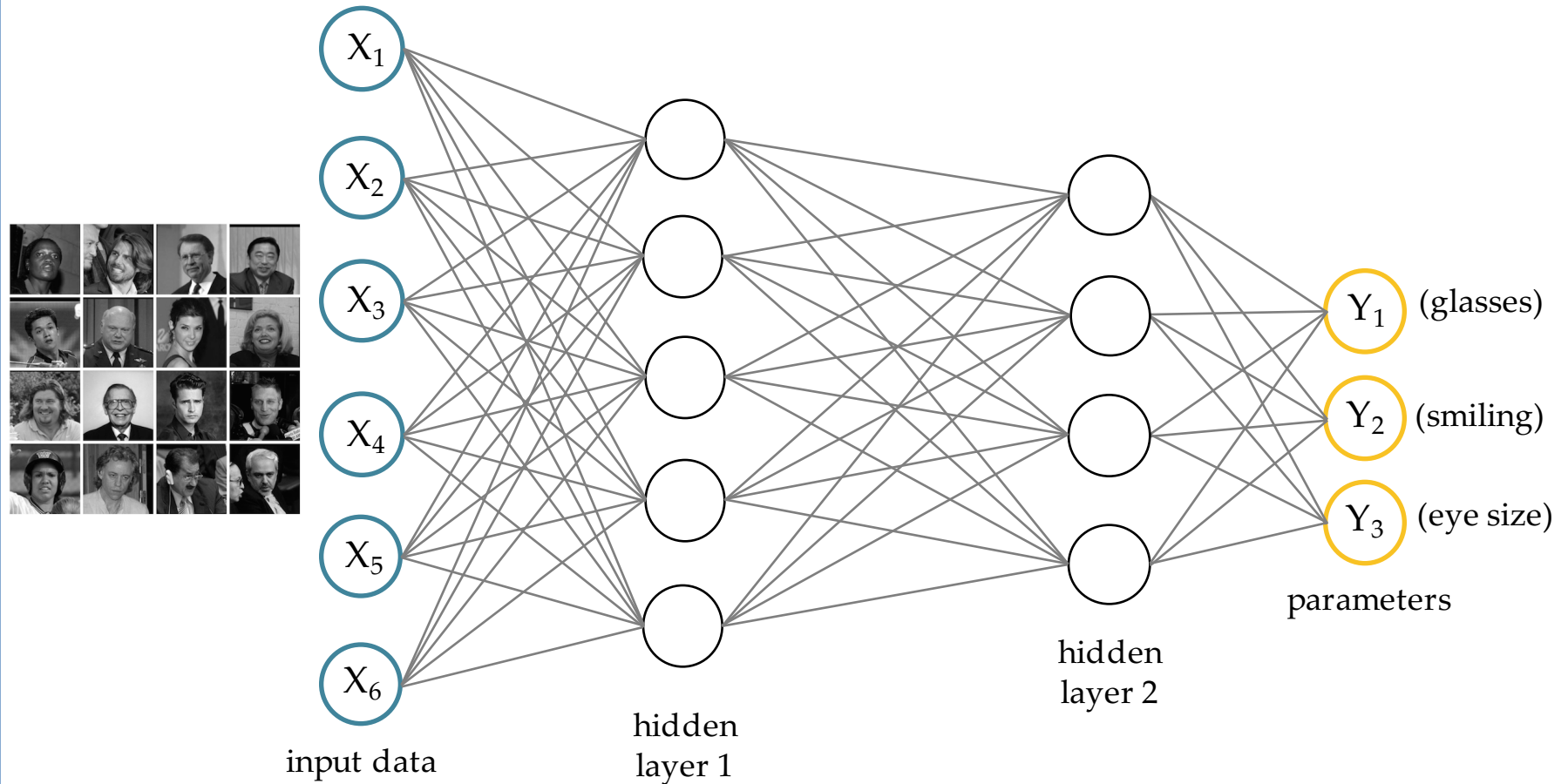
Goal: learn from complicated inputs



Idea: transform data into lower dimension



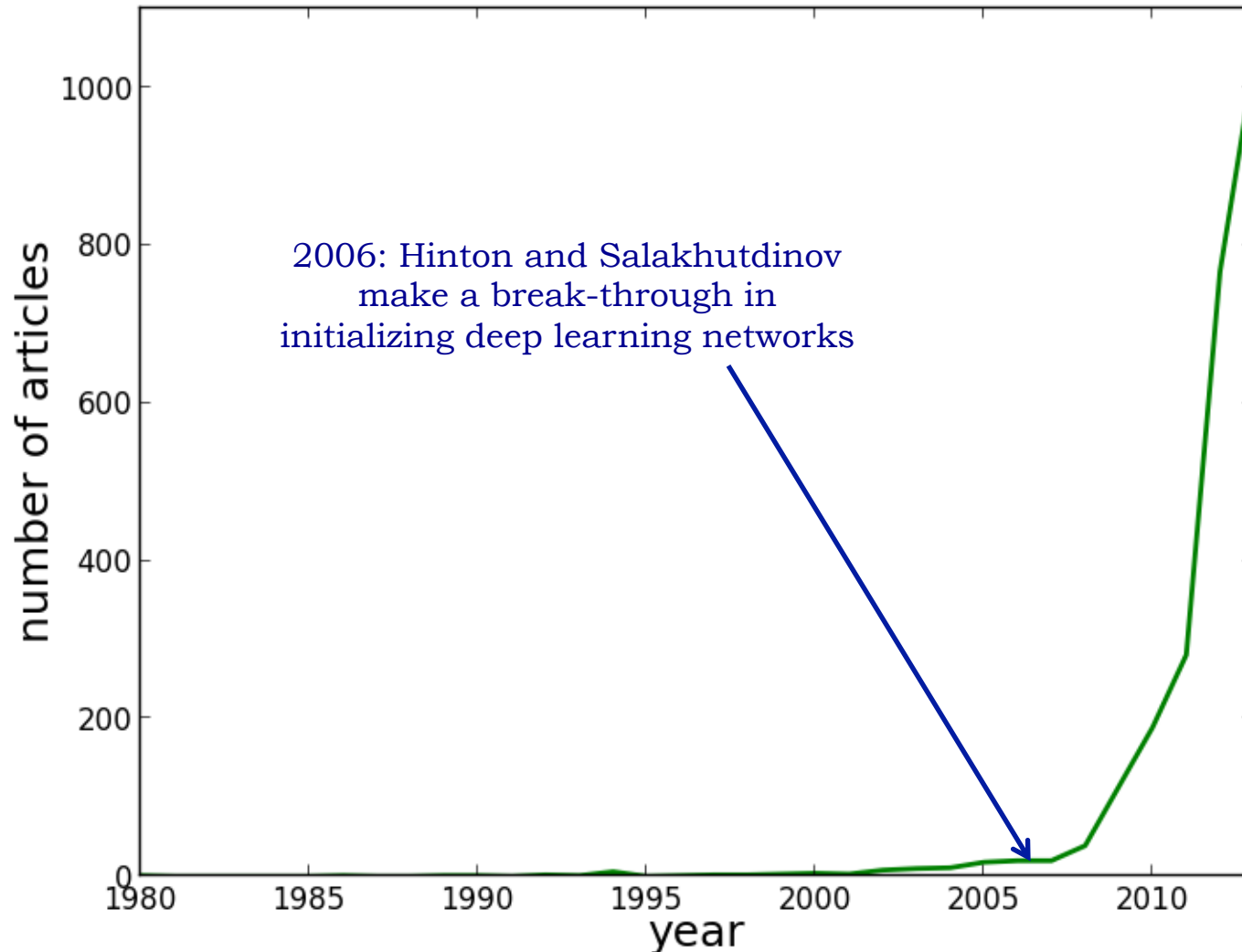
Multi-layer networks = “deep learning”



History of Neural Networks

- Perceptron can be interpreted as a simple neural network
- Misconceptions about the weaknesses of perceptrons contributed to declining funding for NN research
- Difficulty of training multi-layer NNs contributed to second setback
- Mid 2000's: breakthroughs in NN training contribute to rise of “deep learning”

Number of papers that mention “deep learning” over time



Backpropagation

- *High-level goal:* we want to know how the output depends on the input

Backpropagation

- *High-level goal:* we want to know how the output depends on the input
- *Issue:* network is very complicated and overall gradient may be difficult to compute

Backpropagation

- *High-level goal:* we want to know how the output depends on the input
- *Issue:* network is very complicated and overall gradient may be difficult to compute
- *Idea:* use the chain rule to compute local gradients throughout the network

Backpropagation

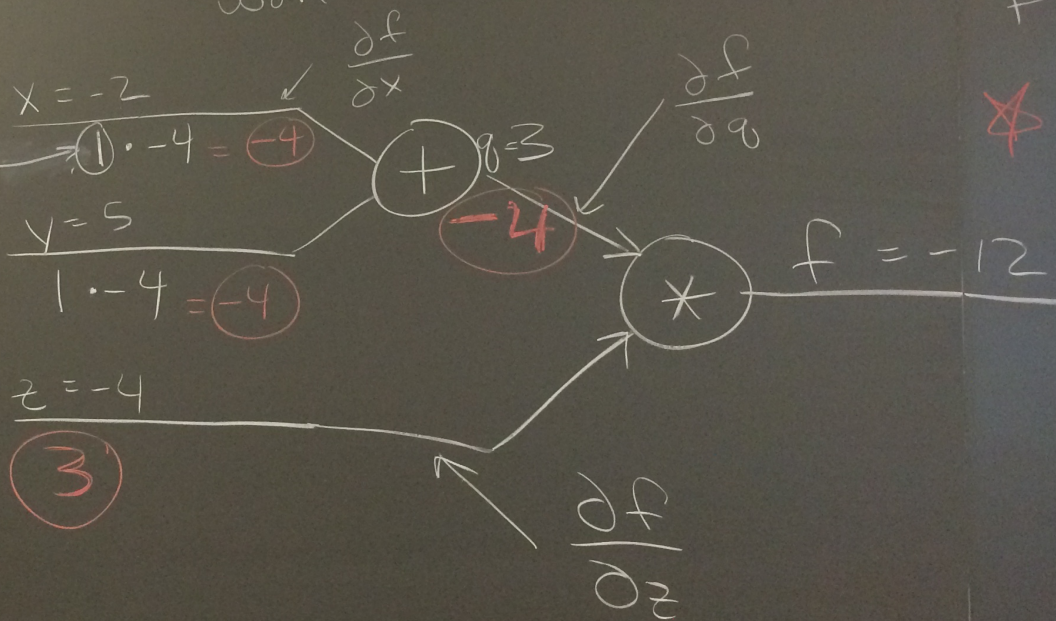
- *High-level goal:* we want to know how the output depends on the input
- *Issue:* network is very complicated and overall gradient may be difficult to compute
- *Idea:* use the chain rule to compute local gradients throughout the network
- *Takeaway:* nodes can know about their value and local gradient without knowing about the network they are imbedded in

$$\frac{\partial q}{\partial x} = 1$$

$$\frac{\partial q}{\partial y} = 1$$

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

want to maximize f



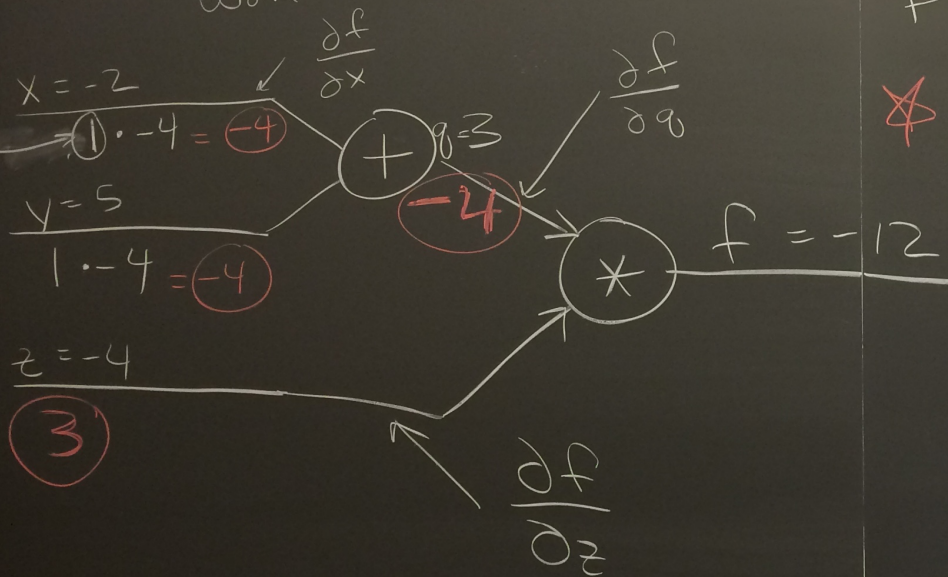
$$q = x + y$$

$$f = qz$$

$$\frac{\partial f}{\partial x} = 1$$

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

want to maximize f



$$q = x + y$$

$$f = q \cdot z$$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial x}$$

chain rule

$$\frac{\partial f}{\partial q} = z$$

$$\frac{\partial f}{\partial z} = q$$