~~Math~~ Computational Foundations for ML

10-60~~6~~7

Geoff Gordon

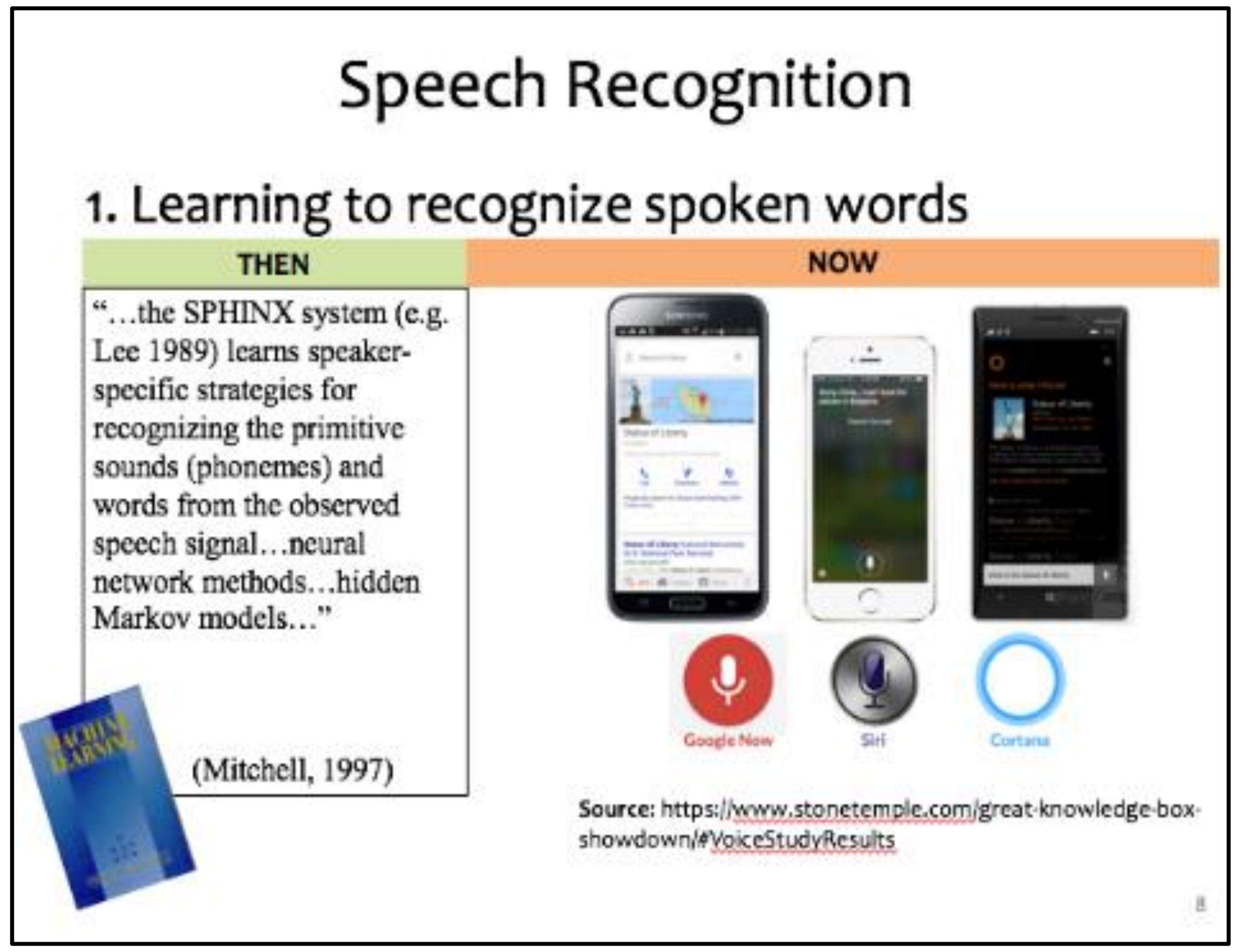
About us

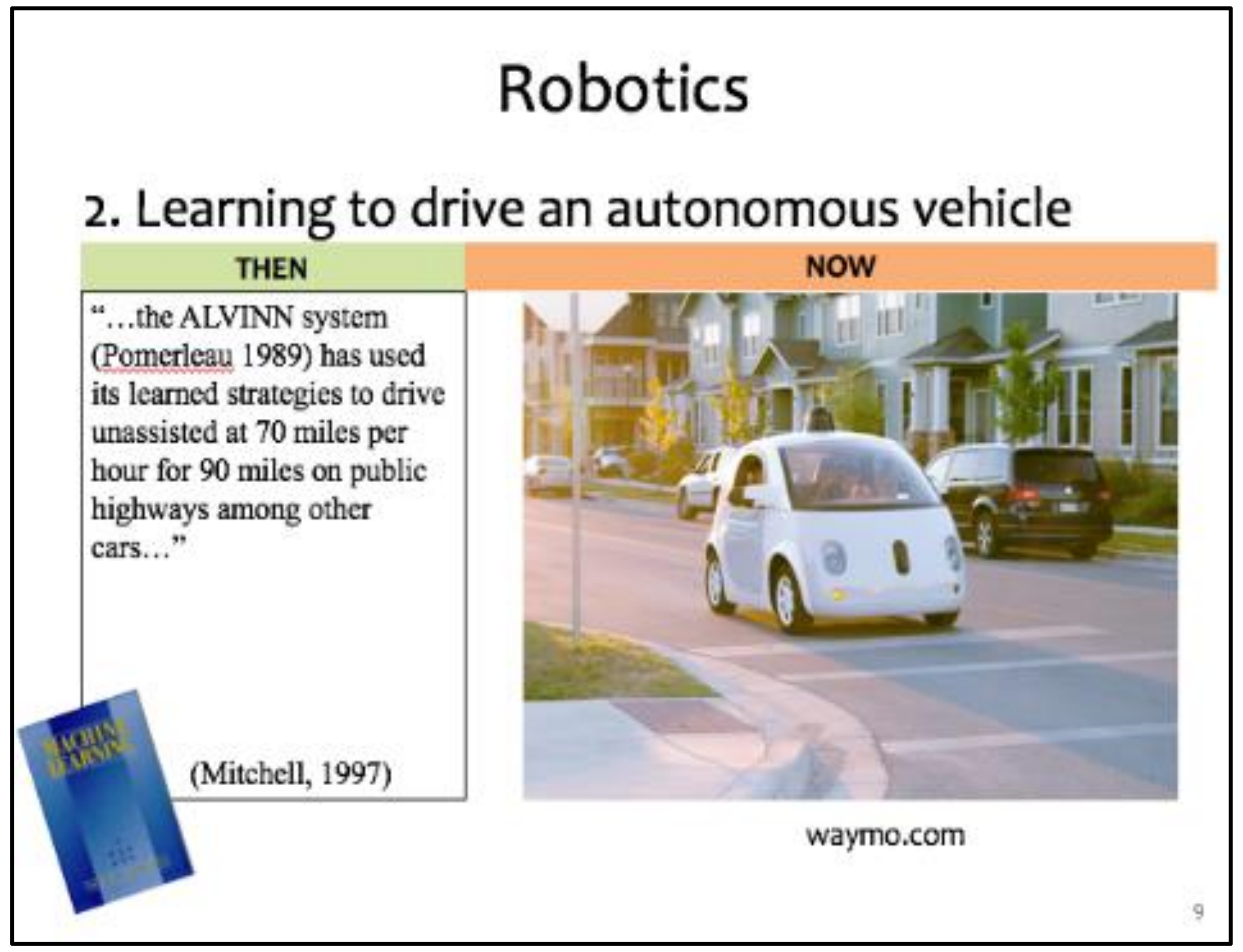
- Me: Geoff Gordon
- TAs:
  - Aditya Paul
  - Aishwarya Jadhav
  - Xiaoyu Xu

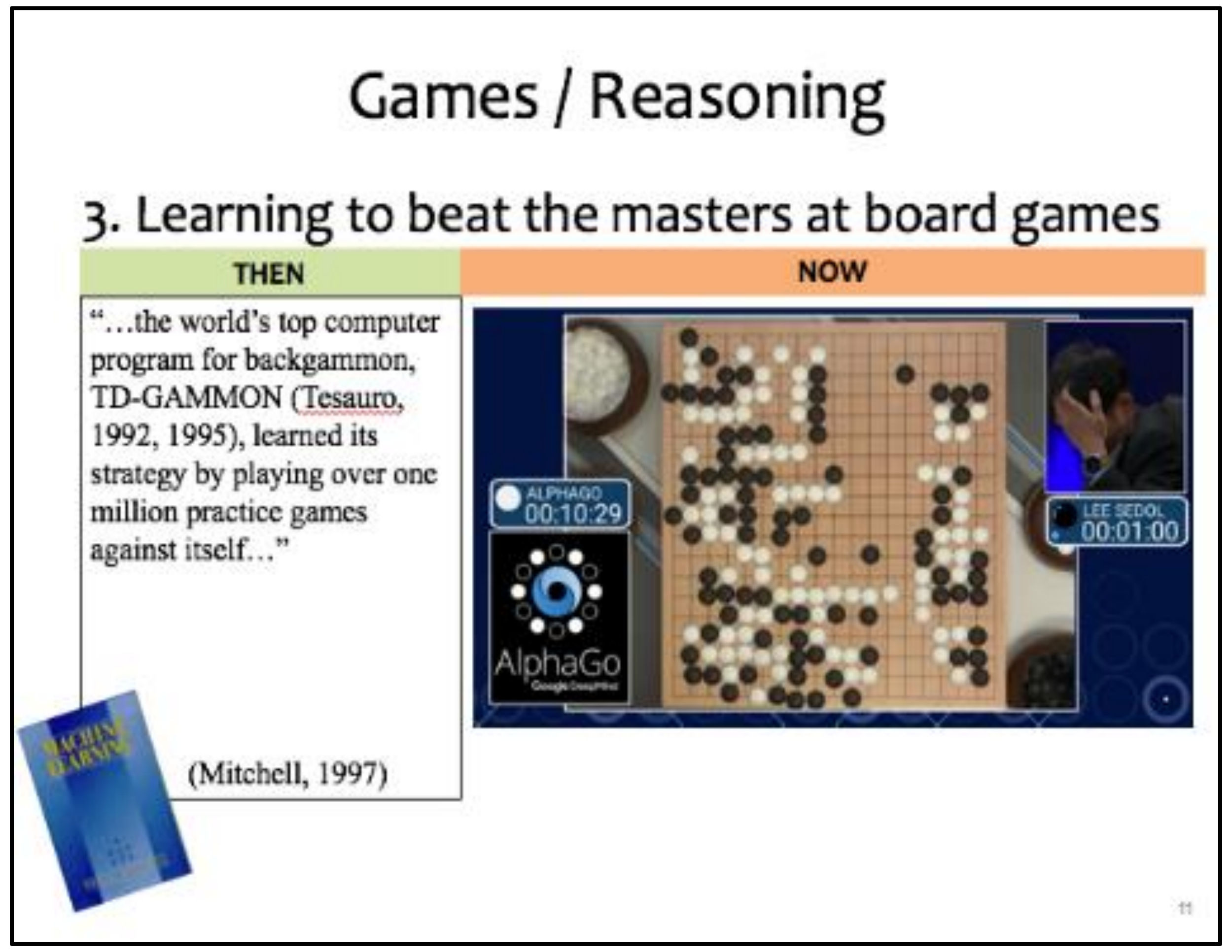
Notes and reminders

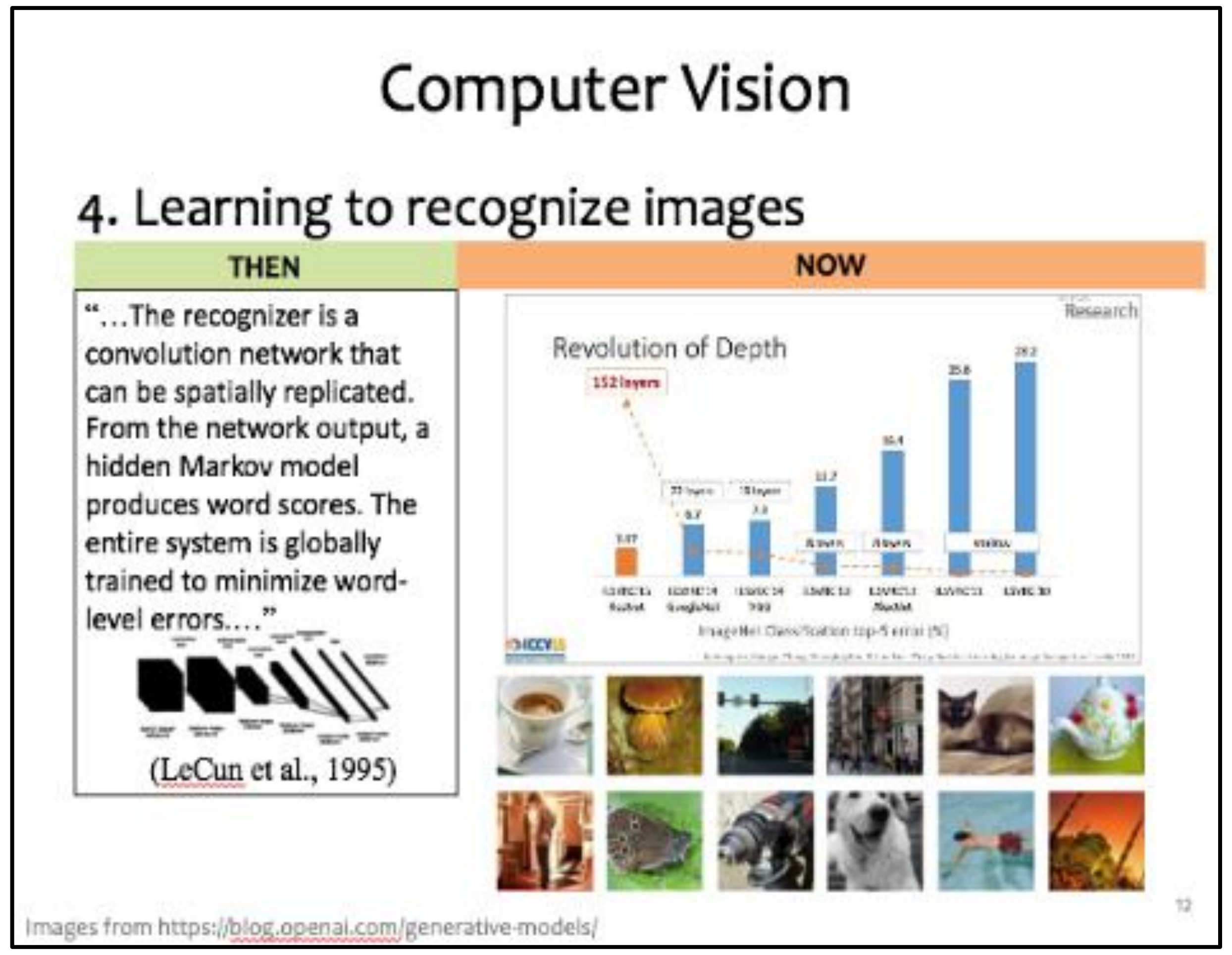
- Location: CUC McKenna (when in person)
- Most weeks: two lectures, one “other”
  - This week: virtual-only lectures today and Wednesday
  - Lab 0 (optional) on Friday: Python review
- https://www.cs.cmu.edu/~ggordon/10606s22/syllabus-and-lecture-outline.html
- Ask questions, participate, help one another learn! https://xkcd.com/1053/

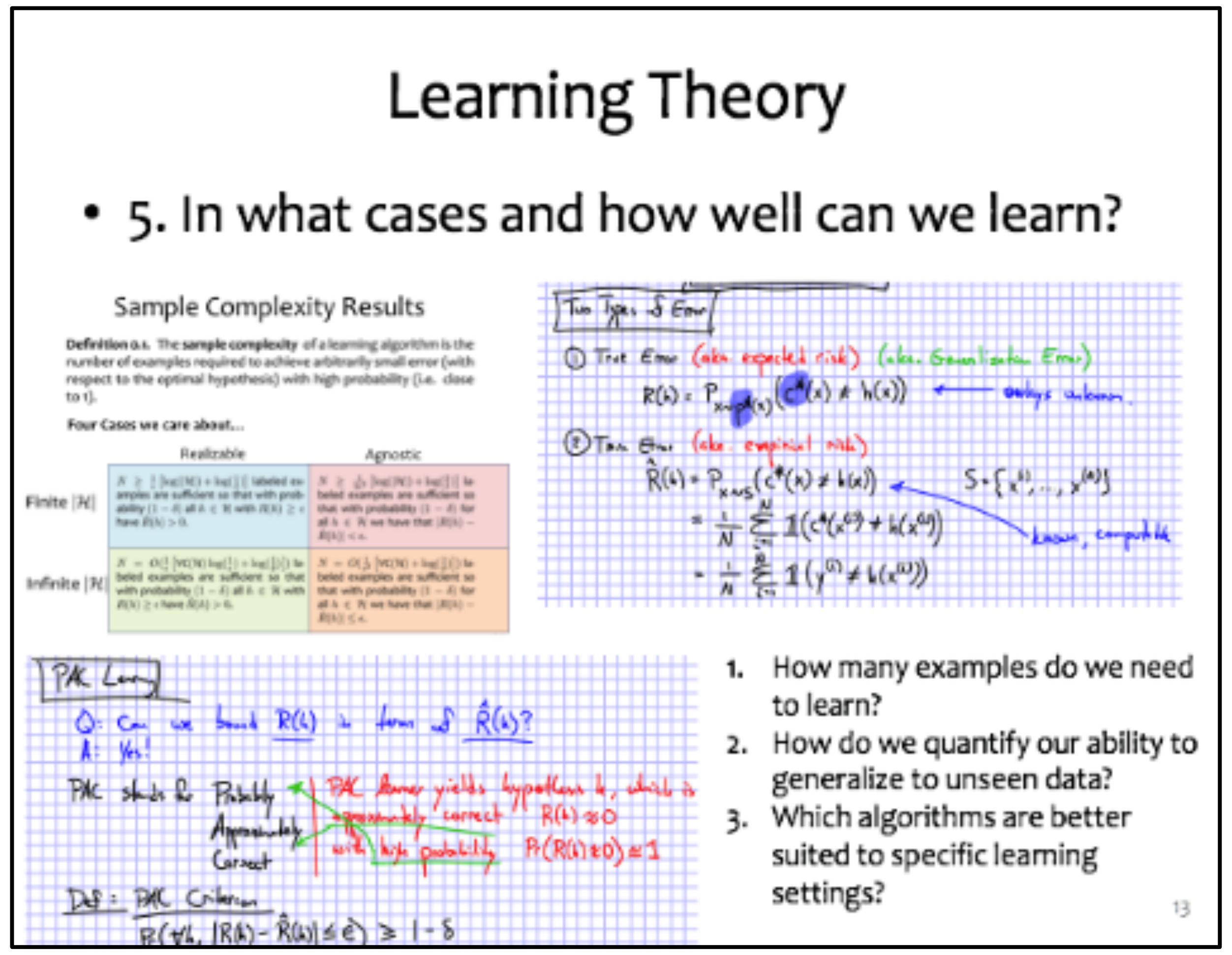
What is ML?











credit: Matt Gormley

- Dual problem (derivation):

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}.\mathbf{w} - \sum_{j=1}^{n} \alpha_j \left[\left(\mathbf{w}.\mathbf{x}_j + b \right) y_j - 1 \right]$$
$$\alpha_j \geq 0, \ \forall j$$

α - weights on training pts (n-dim problem)

Why this course?

Dual SVM – linearly separable case

max $\alpha_j \geq 0$ $d(\alpha)$

- Dual problem (derivation):

$$\max_\alpha \min_{\mathbf{w},b} L(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}.\mathbf{w} - \sum_j \alpha_j \left[\left(\mathbf{w}.\mathbf{x}_j + b \right) y_j - 1 \right]$$
$$\alpha_j \geq 0, \ \forall j$$

$$\frac{\partial L}{\partial w} = w - \sum_j \alpha_j x_j y_j$$

$$\to \frac{\partial L}{\partial \mathbf{w}} = 0 \qquad \Rightarrow \mathbf{w} = \sum_j \alpha_j y_j \mathbf{x}_j$$

$$\to \frac{\partial L}{\partial b} = 0 \qquad \Rightarrow \sum_j \alpha_j y_j = 0$$

$$\frac{\partial L}{\partial b} = \sum_j \alpha_j y_j$$

Dual SVM – linearly separable case

$$\sum_j \alpha_j b y_j = b \sum_j \alpha_j y_j = 0$$

- Dual problem:

$$\max_\alpha \min_{\mathbf{w},b} L(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}.\mathbf{w} - \sum_j \alpha_j \left[\left(\mathbf{w}.\mathbf{x}_j + b \right) y_j - 1 \right]$$
$$\alpha_j \geq 0, \ \forall j$$

$$\Rightarrow \mathbf{w} = \sum_j \alpha_j y_j \mathbf{x}_j \qquad \Rightarrow \sum_j \alpha_j y_j = 0$$

$$\underbrace{\frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j x_i \cdot x_j}_{} - \underbrace{\sum_j \alpha_j \left(\sum_i \alpha_i y_i x_i \cdot x_j \right) y_j}_{} + \sum_j \alpha_j$$

$$\sum_j \alpha_j - \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j x_i \cdot x_j =: d(\alpha)$$

This course is a bit odd

- Most courses:
  - teach one semester’s worth of material in one semester
  - go over enough examples of every topic that you can learn it from scratch with no trouble
- This course:
  - several semester’s worth of material in one!
  - assumption: you’ve seen at least some of it before, need to work on rest
  - not enough examples on any one topic to learn from scratch: you must ask questions and seek out your own material
- Benefit: we cover a lot of ground, help build a strong base for ML courses

Formal systems

Objects $\in$ Data type

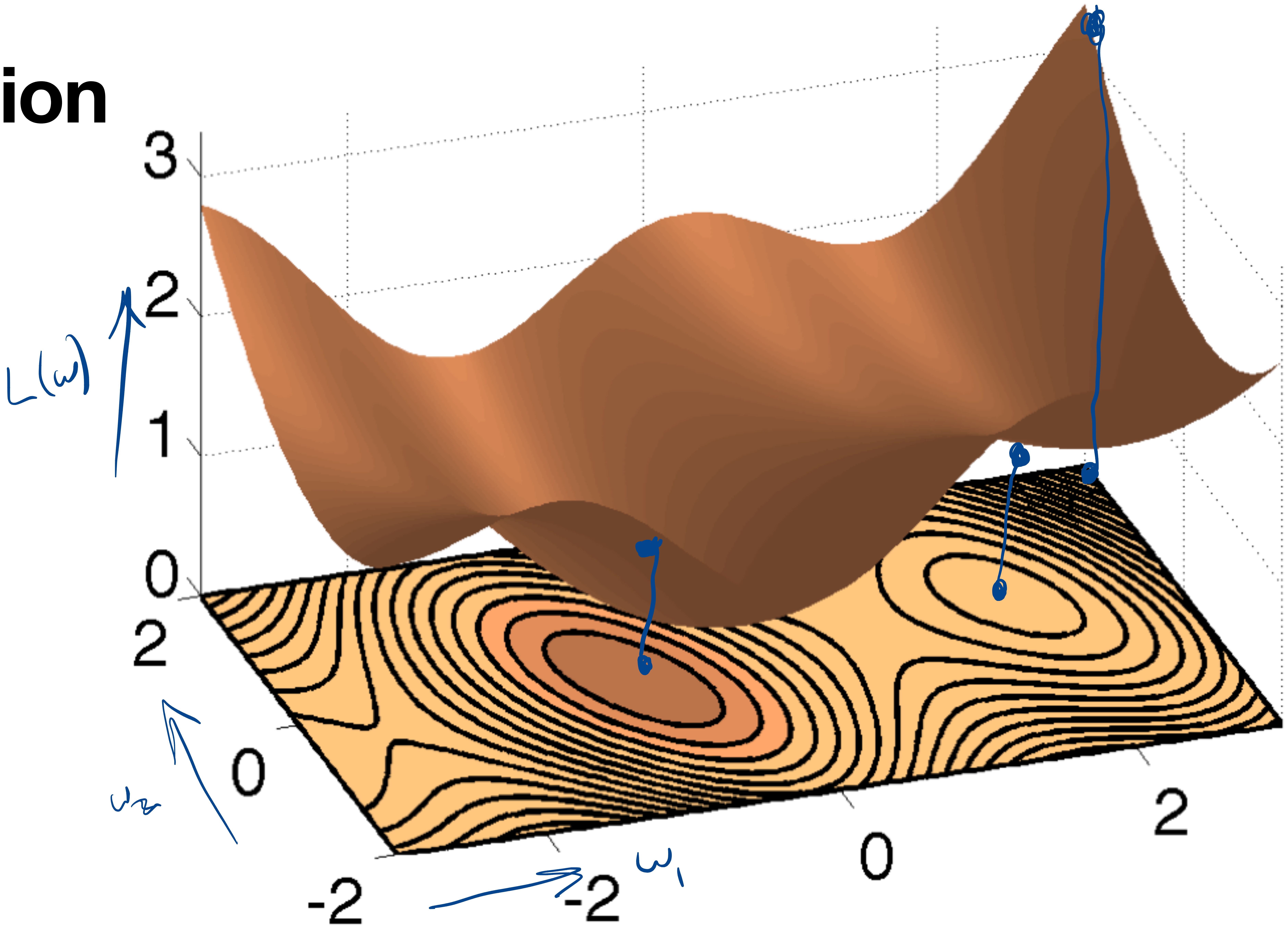
↳ API

Optimization

$\min_w L(w)$

$\uparrow$

$\in \mathbb{R}^2$



Optimization from data

$P(x_i \mid w)$

$X \quad x_1, x_2 \; \ldots \; x_T$

$P(w \mid X) \propto P(x \mid w) \cancel{P(w)} / \cancel{P(x)}$

$L(w)$

$\downarrow$

$\max_w P(w \mid X)$

$\hookrightarrow P(x_1 \mid w) \, P(x_2 \mid w) \ldots P(x_T \mid w)$

$P(y_i \mid x_i, w)$

$P(w \mid X, Y) \propto P(Y \mid X, w) \cancel{P(w \mid X)}$

$P(y_1 \mid x_1, w) \, P(y_2 \mid x_2 \mid w) \ldots P(y_T \mid x_T, w) \quad \cancel{P(Y \mid X)}$

Example: classification

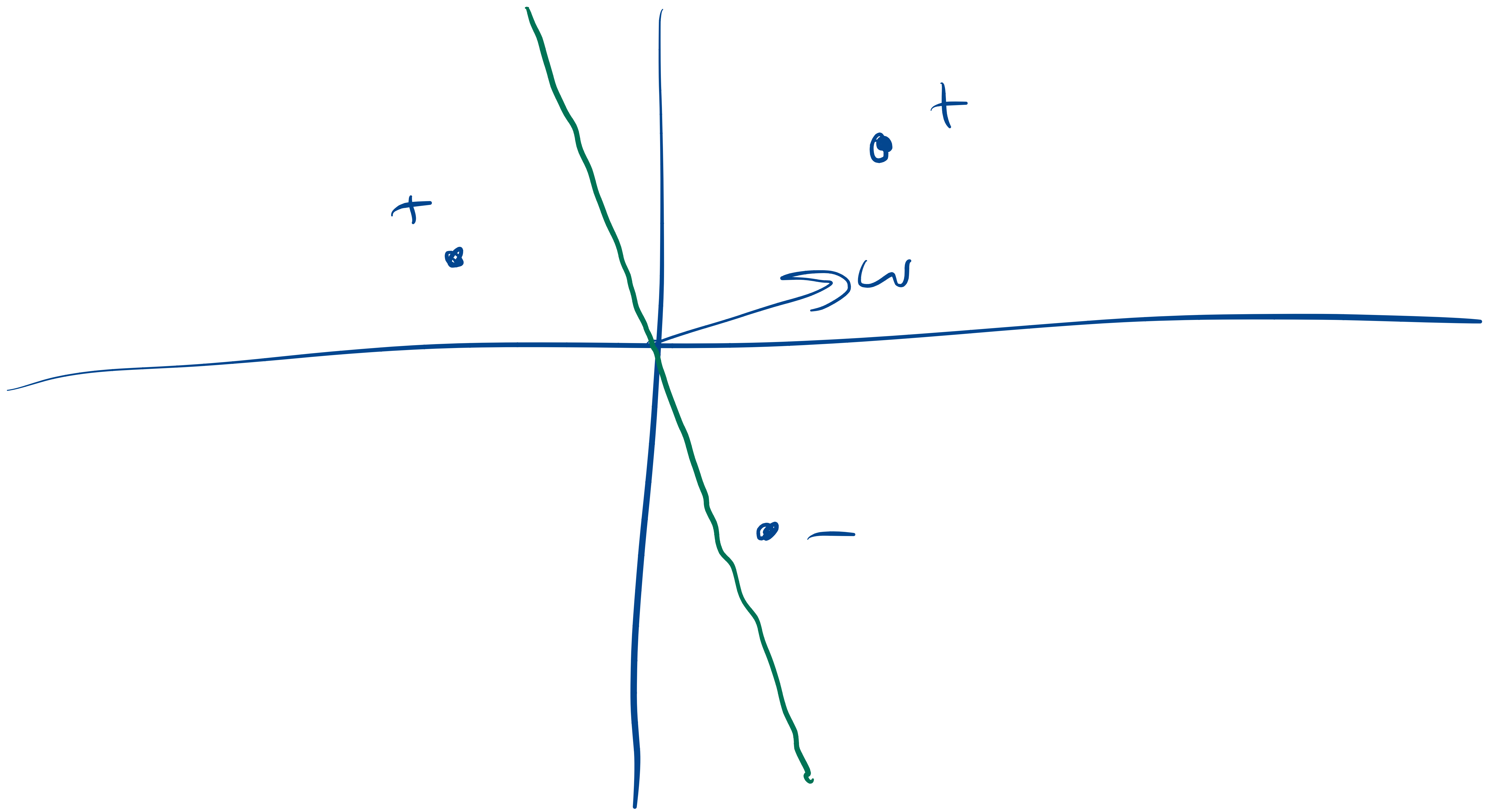
$$\left\{ \begin{array}{l} x_i^1 \to \mathbb{R} \\ x_i^2 \to \mathbb{R} \\ y_i \to \pm 1 \end{array} \right.$$

$$L(w) = \prod_i P(y_i \mid x_i, w)$$

Classifier:

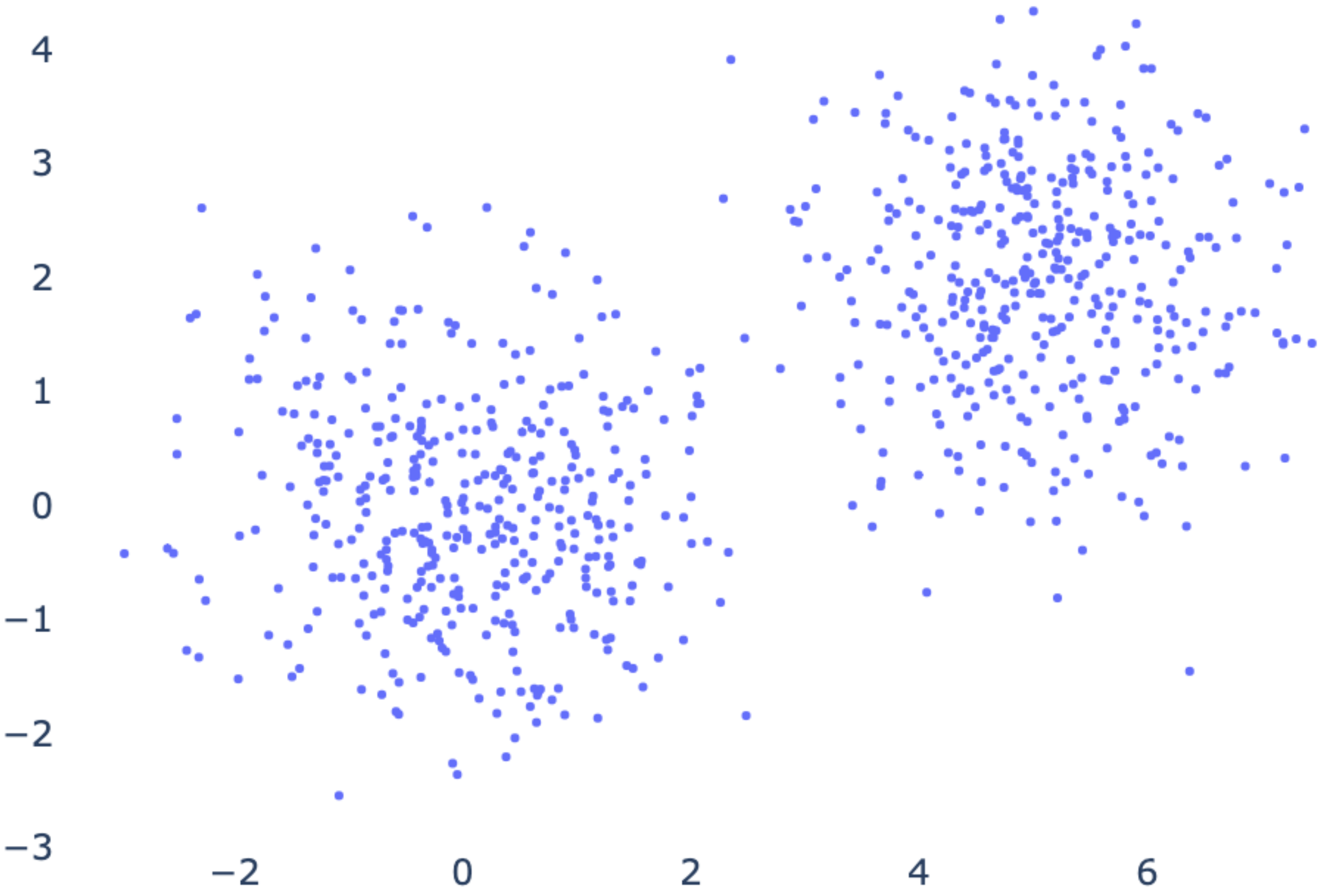
input $x \in \mathbb{R}^k$

output $y \in \{\pm 1\}$

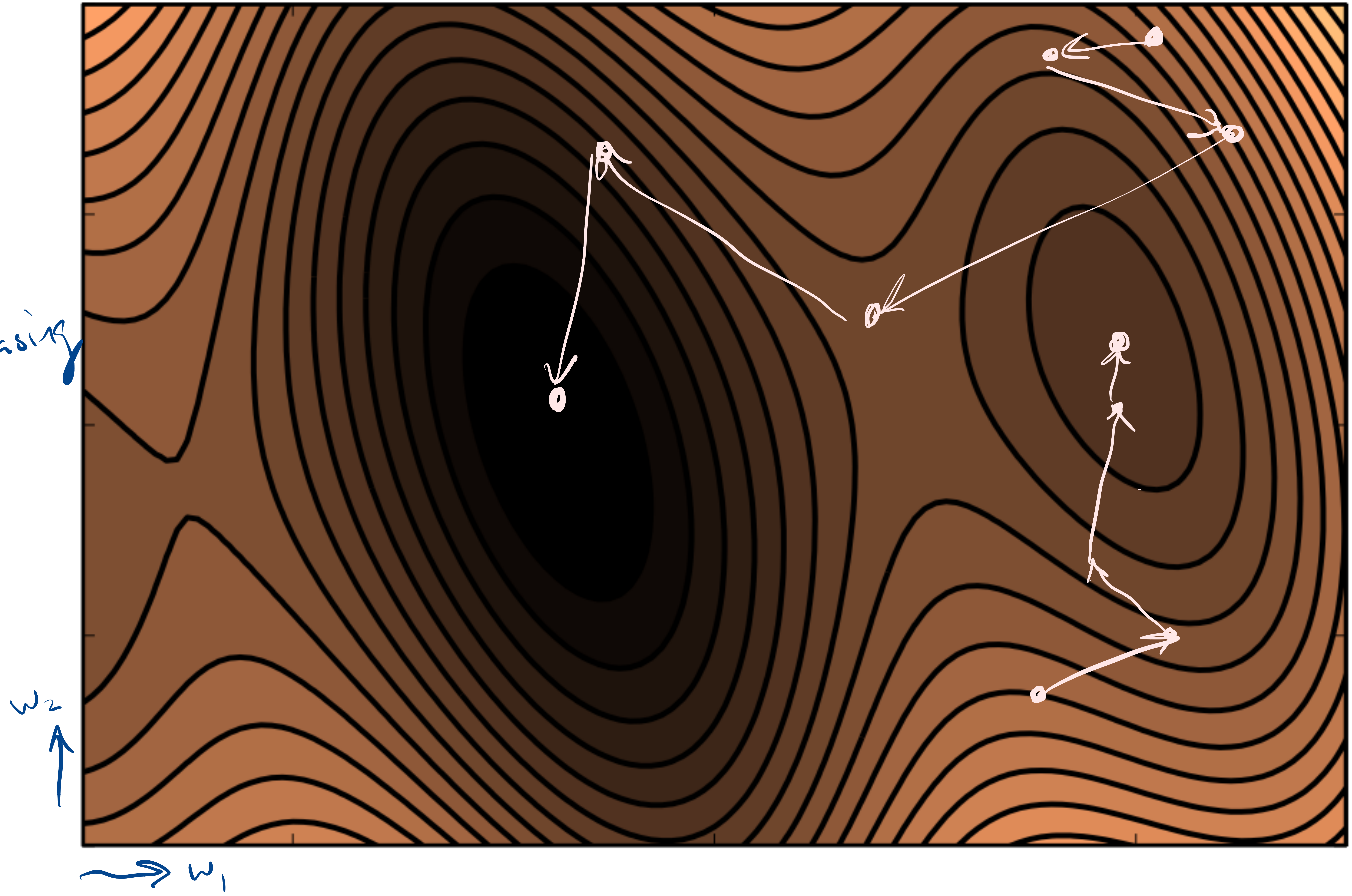


Example: density estimation

$P(w) \propto \prod_i P(x_i | w)$



Optimizer: tries to follow steps of decreasing $L(w)$



Gradient descent

start at $w^{(0)}$

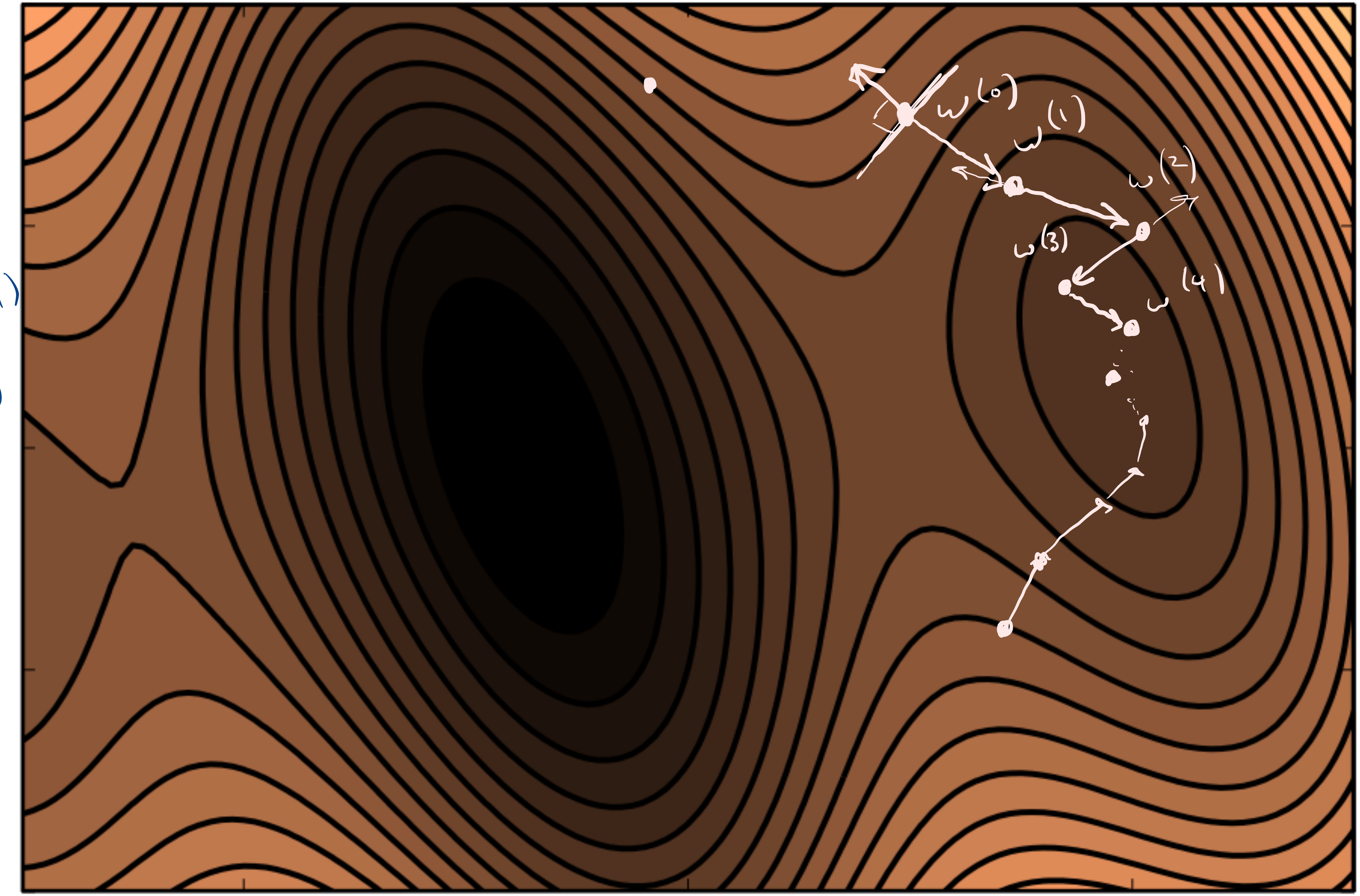
repeat $i = 1, 2, \ldots$

$w^{(i)} = w^{(i-1)} - \eta^{(i)} g^{(i-1)}$

$g^{(i-1)} = \nabla L(w^{(i-1)})$

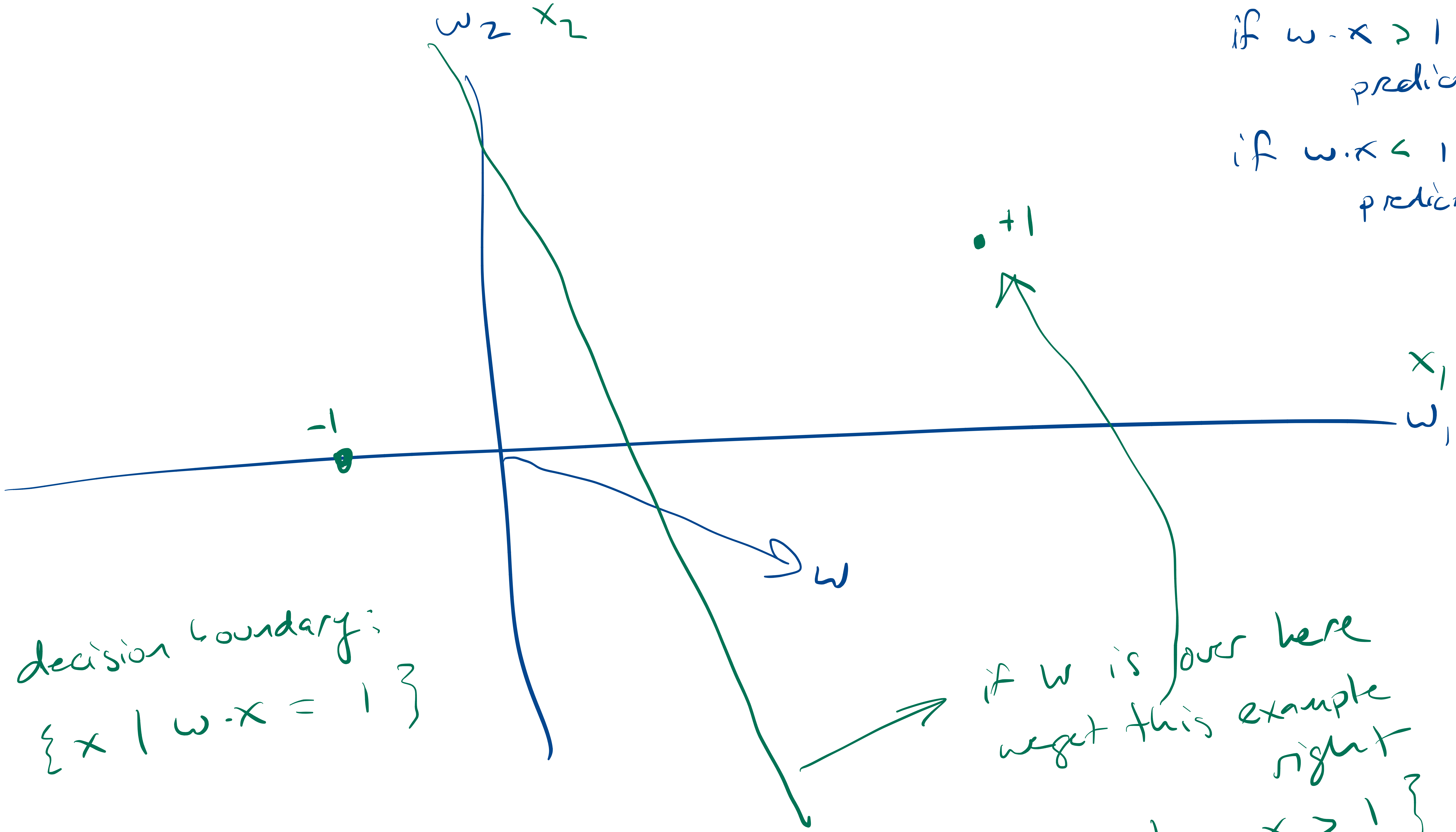
$\eta^{(i)} \in \mathbb{R}$

$\eta^{(i)} > 0$



if $w \cdot x > 1$
predict $+1$

if $w \cdot x < 1$
predict -1



decision boundary:
$\{x \mid w \cdot x = 1\}$

if w is over here
we get this example
right

$\{w \mid w \cdot x > 1\}$

Example

go to repl.it