Important announcements

Midterm details:

- 25 multiple choice (select all that apply) questions using scantron (Bring a pencil!!)
- Covers up to (and including) today lecture up to (and including) sec 6.2 in the course notes
- Formula sheet will be provided (no cheat sheet) will post tonight
- You can bring a calculator
- Similar to assignment questions and examples in course notes
- we will use lecture on Thursday to do a review session for the midterm exam.

=> Our goal (from supervised learning lecture)

Defining $\mathcal{A}(D)$: Empirical Risk Minimization (ERM)

Estimation:

Use D to estimate $L(f)$ for all $f \in \mathcal{F} \subset \{f | f: \mathcal{X} \to \mathcal{Y}\}$

call the estimate $\hat{L}(f)$

Optimization:

pick $\hat{f}$ to be the $f \in \mathcal{F}$ that minimizes $\hat{L}(f)$

$\mathcal{F}$: Function class

Optimization

finding the best solution from a set of possible solutions

Usually this means finding the minimum or maximum value of some function

we will care about:

$$\min_{w \in \mathcal{W}} g(w)$$

minimum value of $g(w)$ over all $w \in \mathcal{W}$

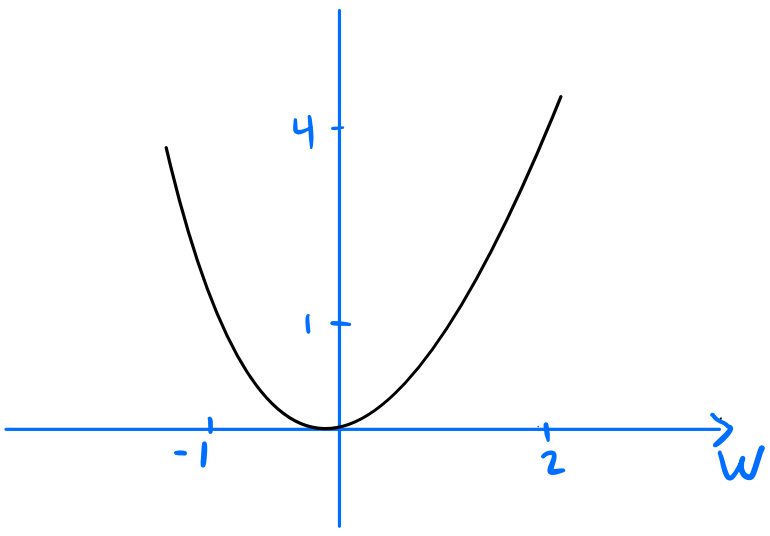
or $w^* = \operatorname{argmin}_{w \in \mathcal{W}} g(w)$

the $w \in \mathcal{W}$ that achieves the minimum value of $g(w)$

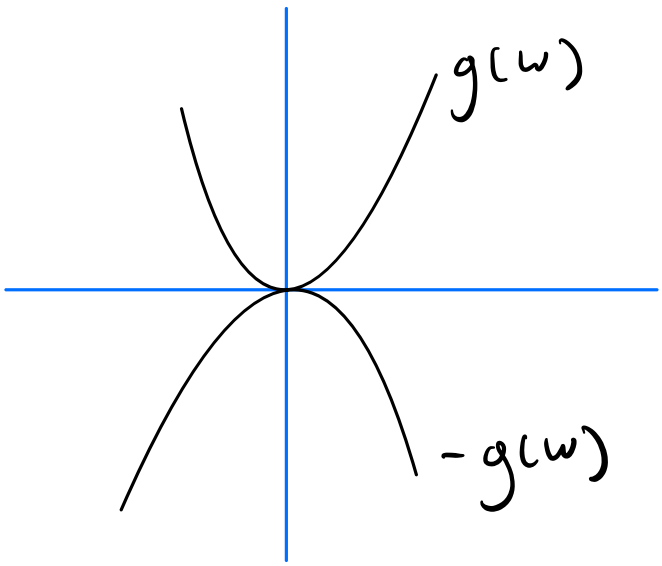
w^* is a "minimizer"

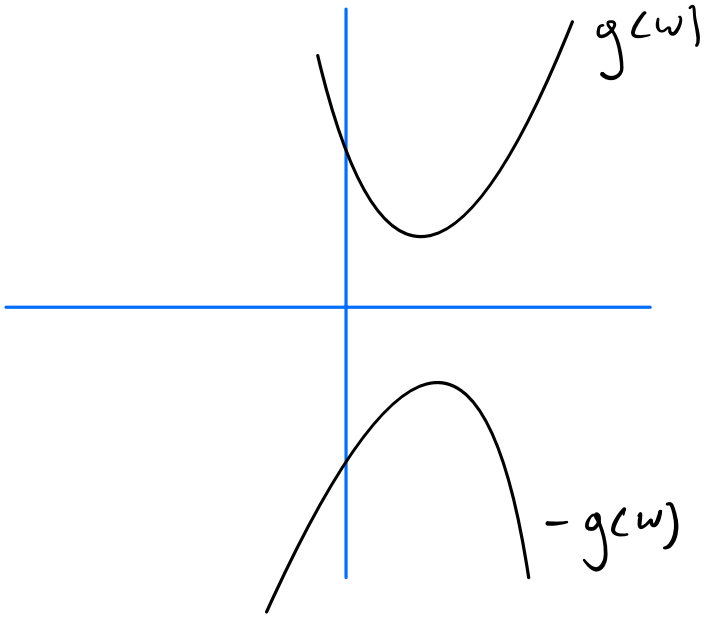
$$\min_{w \in \mathcal{W}} g(w) = g(w^*)$$

Ex: $g(w) = w^2$



Note: There is a relationship between minimizing and maximizing





How do we solve minimization problems?

Cases:

1. If W is discrete

2. If W is continuous

Additional assumptions:

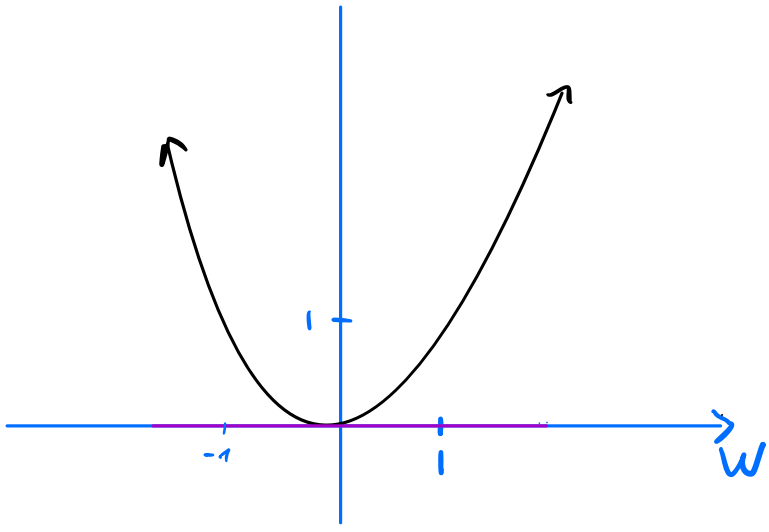
Cases: 1.

2.

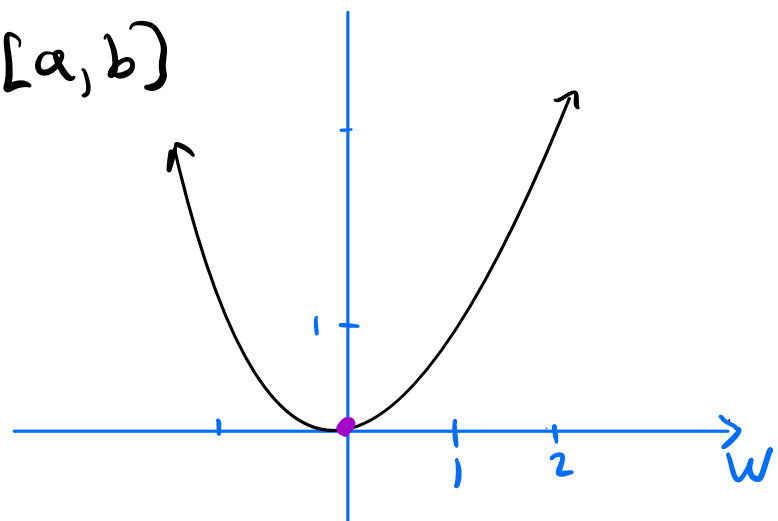
Twice doffentiable:

Convex:

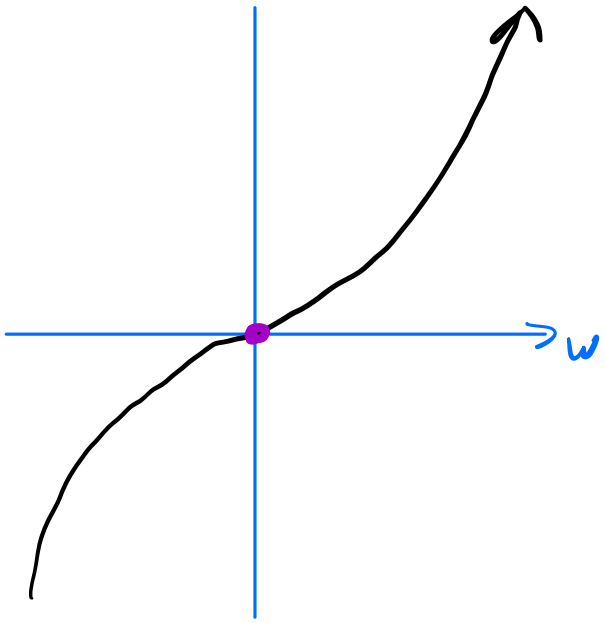
Ex: $g(w) = w^2$, $\mathcal{W} = \mathbb{R}$



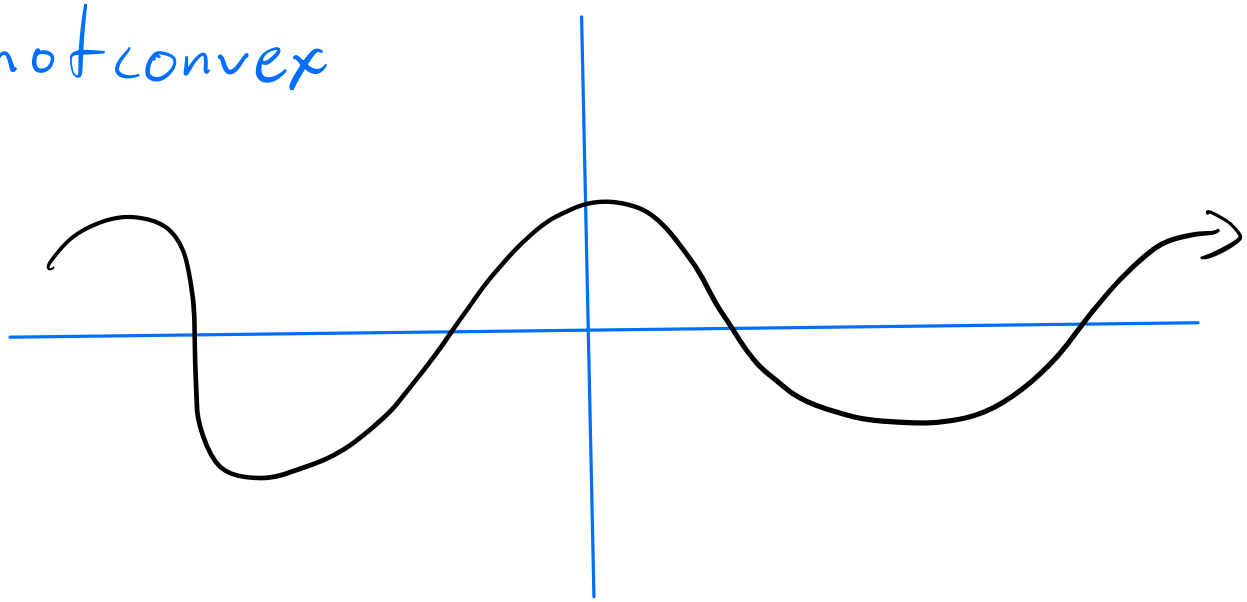
Ex: $g(w) = w^2$, $\mathcal{W} = [1, 2] = [a, b]$



Ex: $g(w) = w^3$, $\mathcal{W} = \mathbb{R}$

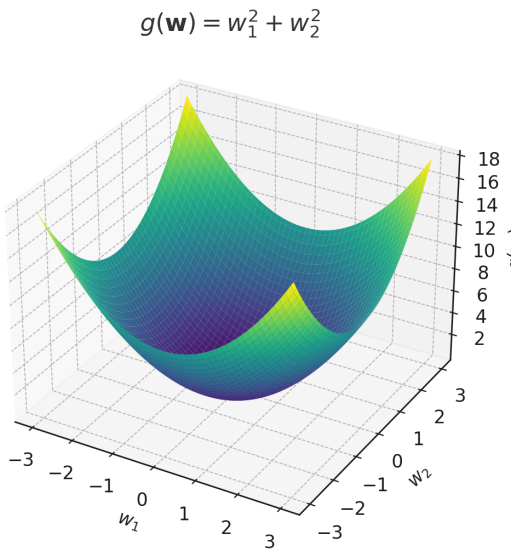


Ex: not convex



Multidimensional Minimization

Ex: $g(\vec{w}) = g(w_1, w_2) = w_1^2 + w_2^2$, $\mathcal{W} = \mathbb{R}^2$



Finding a good predictor (Linear Regression)

Optimization step of ERM