Optimization

finding the best solution from a set of options

Usually this means minimizing/maximizing some function

$$g : \mathcal{W} \to \mathbb{R}$$

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

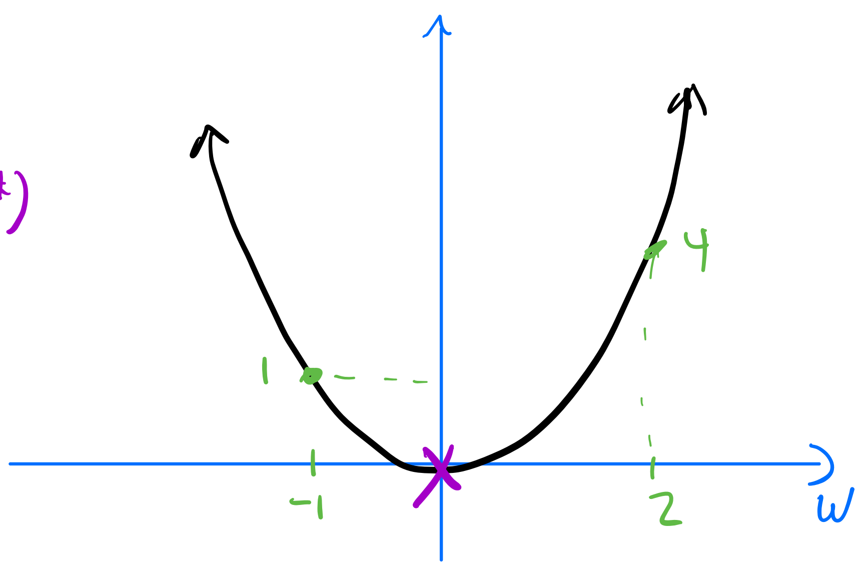
$$\arg\min_{w \in \mathcal{W}} g(w) = w^*$$

<u>Ex</u>: $g(w) = w^2$, $g : \mathcal{W} \to \mathbb{R}$

$\mathcal{W} = \mathbb{R}$

$$\min_{w \in \mathbb{R}} g(w) = 0 = g(w^*)$$

$$\arg\min_{w \in \mathbb{R}} g(w) = 0 = w^*$$

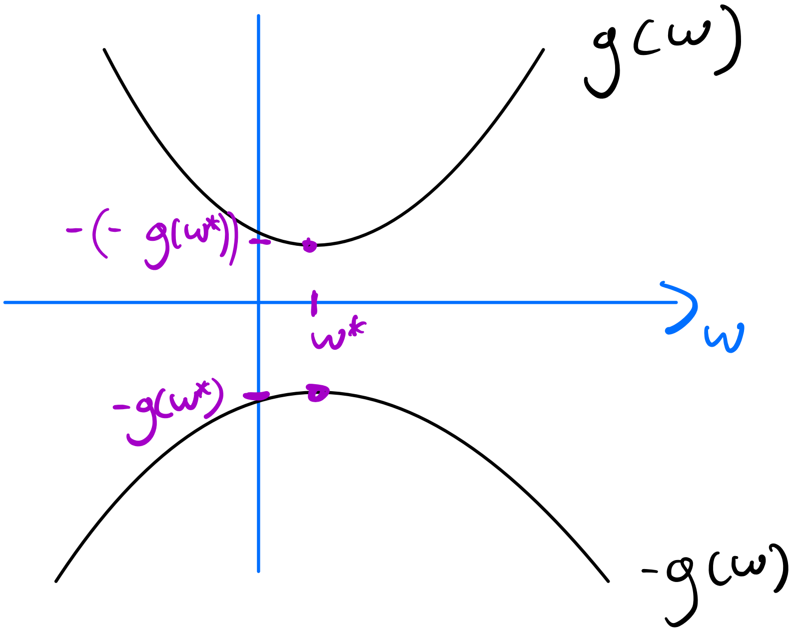


$\mathcal{W} = \{-1, 2\}$

$\min_{w \in \mathcal{W}} g(w) = 1$ $\quad$ $\underset{w \in \mathcal{W}}{\operatorname{argmin}}\, g(w) = -1$

$= g(w^*)$

$= (-1)^2$



$$w^* = \underset{w \in \mathcal{W}}{\operatorname{argmin}}\, g(w) = \underset{w \in \mathcal{W}}{\operatorname{argmax}}\, -g(w)$$

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

How to solve minimization problems?

Cases:

1. If $\mathcal{W}$ is discrete we compare $g(w)$ for $w \in \mathcal{W}$

2. If $\mathcal{W}$ is continuous we can sometimes use derivatives

Assumptions: $\mathcal{W}$ is continuous

g is convex and twice differentiable

Twice differentiable: $g''(w)$ exists for $w \in \mathcal{W}$

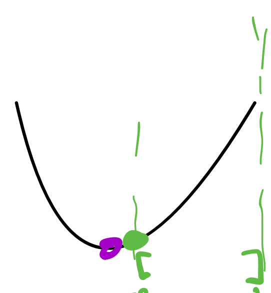
Convex: $g''(w) \geq 0$ for all $w \in \mathcal{W}$

Ex: $g(w) = w^2$, $g'(w) = 2w$, $g''(w) = 2 \geq 0$

$\mathcal{W} = \mathbb{R} / [-1, 1]$

$g'(w) = 2w = 0$

$\Rightarrow w^* = 0$



$\mathcal{W} = [1, 2]$

$g(1) = 1 \Rightarrow w^* = 1$

$g(2) = 4$

Finding the minimum:

1. if $\mathcal{W} = \mathbb{R}$ then w^* is the soln to $g'(w) = 0$

2. if $\mathcal{W} = [a, b]$ then w^* is the the soln of $g'(w) = 0$ if it is in $\mathcal{W}$

   otherwise w^* is a or b

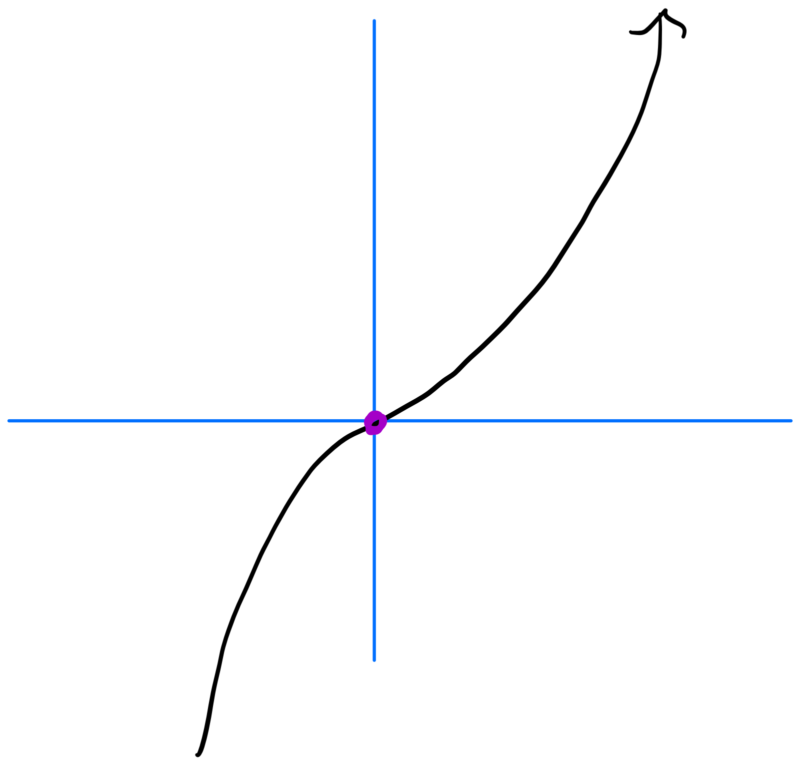
Ex: $g(w) = w^3$, $g'(w) = 3w^2$, $g''(w) = 6w \not\geq 0$?

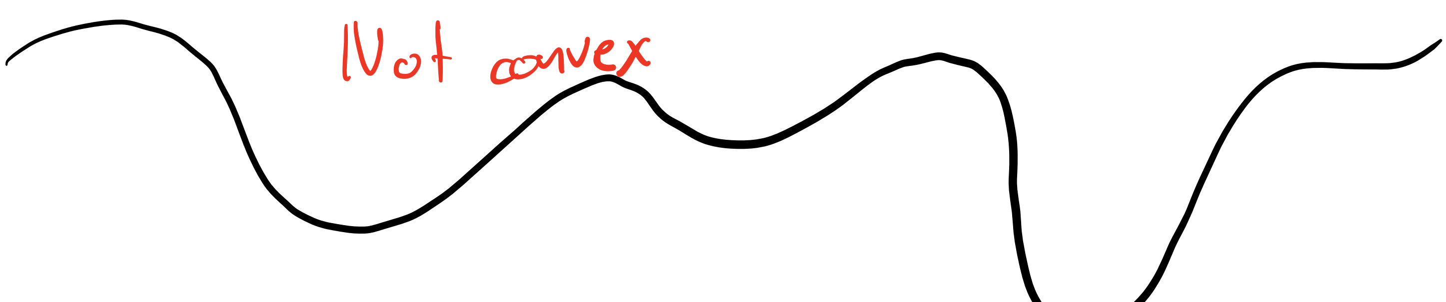
if $w = -1$

$g'(w) = 3w^2 = 0$

$\Rightarrow w^* = 0$

$\mathcal{W} = \mathbb{R}$





Multidimensional minimization

if $\mathcal{W} = \mathbb{R}^d$ $\quad d > 1 \quad$ and $\quad g(\vec{w})$ is convex

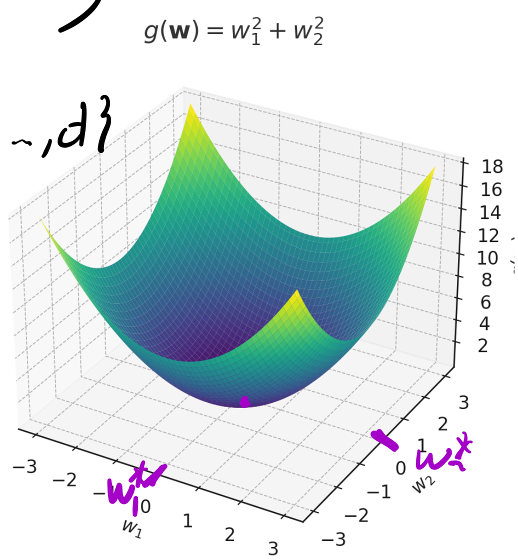
Note: if $d > 1$ I will always Tell you if its convex

$$\boxed{A \succeq 0}$$

$$(w_1^*, \ldots, w_d^*)^T = \vec{w}^* = \underset{\vec{w} \in \mathcal{W} = \mathbb{R}^d}{\operatorname{argmin}} \; g(\vec{w})$$

We calculate $\vec{w}^*$ by solving

$$\frac{\partial g}{\partial w_j}(\vec{w}) = 0 \quad \text{for all } j \in \{1, \ldots, d\}$$



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

$\mathcal{W} = \mathbb{R}^2$

$$\frac{\partial g}{\partial w_1} = 2w_1 = 0 \Rightarrow w_1^* = 0$$

$$w_2^* = 0$$

$$\vec{w}^* = (0, 0)^T$$

$g(w_1, w_2) = w_1^2 w_2^2$

$$\frac{\partial g}{\partial w_1}(w_1, w_2) = 2w_1 w_2^2 = 0 \quad \Rightarrow w_1^* = 0$$

$$\frac{\partial g}{\partial w_2}(w_1, w_2) = 2w_2 w_1^2 = 0 \quad \Rightarrow w_2^* = 0$$

ERM for Regression with linear functions

$\mathcal{X} = \mathbb{R}^{d+1}$; $\mathcal{Y} = \mathbb{R}$

$\mathcal{F} \subseteq \{f \mid f : \mathbb{R}^{d+1} \to \mathbb{R}\}$

$\mathcal{D} = ((\vec{x}_1, \vec{y}_1), \ldots, (\vec{x}_n, \vec{y}_n))$

$$\hat{L}(f) = \frac{1}{n} \sum_{i=1}^{n} \ell(f(\vec{x}_i), y_i)$$

we want $\boxed{\hat{f} = \underset{f \in \mathcal{F}}{\arg\min}\ \hat{L}(f)}$

$\mathcal{F} = \{f \mid f : \mathbb{R}^{d+1} \to \mathbb{R}$ and $f(\vec{x}) = \vec{x}^T \vec{w}$ for some $\vec{w} \in \mathbb{R}^{d+1}\}$

every $f \in \mathcal{F}$ has a $\vec{w} \in \mathbb{R}^{d+1}$ associated

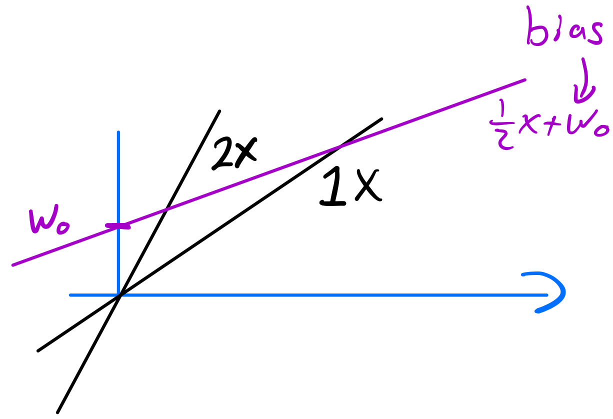
$$\min_{f\in\mathcal{F}} \hat{L}(f) = \min_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \ell(f(\vec{x}_i), y_i)$$

parameters or weights

$$= \min_{\vec{w}\in\mathbb{R}^{d+1}} \frac{1}{n}\sum_{i=1}^{n} \ell(\vec{x}_i^T\vec{w}, y_i)$$

$\mathcal{A}(D) = \hat{f}$ such that

$\hat{f}(\vec{x}) = \vec{x}^T\hat{\vec{w}}$



$$\hat{\vec{w}} = \arg\min_{\vec{w}\in\mathbb{R}^{d+1}} \underbrace{\frac{1}{n}\sum_{i=1}^{n} \ell(\vec{x}_i^T\vec{w}, y_i)}_{}$$

$$= \hat{L}(\vec{w}) = g(\vec{w})$$

is this convex?

yes if ℓ is squared loss

$$f(\vec{x}) = \vec{x}^T\vec{w} + w_0 = 1\cdot w_0 + x_1 w_1 + \dots + x_d w_d$$

$$= (x_0 = 1, x_1, \dots, x_d)(w_0, w_1, \dots, w_d)^T$$

Solve for $\hat{\vec{w}}$

$$\hat{L}(\vec{w}) = \frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i^T\vec{w} - y_i)^2$$

$$= \frac{1}{n}\left[(\vec{x}_1^T\vec{w} - y_1)^2 + \dots + (\vec{x}_n\vec{w} - y_n)^2\right]$$

$$\hat{L}_i(\vec{w}) = (\vec{x}_i^T\vec{w} - y_i)^2 = (x_{i0}w_0 + \dots + x_{ij}w_j + \dots + x_{id}w_d - y_i)^2$$

$$= \frac{1}{n}\left[\hat{L}_1(\vec{w}) + \dots + \hat{L}_n(\vec{w})\right]$$

$$\frac{\partial \hat{L}}{\partial w_j} = \frac{1}{n}\left[\frac{\partial \hat{L}_1}{\partial w_j} + \dots + \frac{\partial \hat{L}_n}{\partial w_j}\right] \qquad j \in \{0, \dots, d\}$$

$$\frac{\partial \hat{L}_i}{\partial w_j}(\vec{w}) = 2(\vec{x}_i^T\vec{w} - y_i)\,x_{ij}$$

$$= \frac{2}{n}\left[x_{1j}(\vec{x}_1^T\vec{w} - y_1) + \dots + x_{nj}(\vec{x}_n^T\vec{w} - y_n)\right]$$

$$= \frac{2}{n}\sum_{i=1}^{n} x_{ij}(\vec{x}_i^T\vec{w} - y_i)$$

$$= 0 \qquad \text{for } j \in \{0, \dots, d\}$$

$$\frac{2}{n}\sum_{i=1}^{n} x_{ij}\left(\vec{x}_i^T\vec{w} - y_i\right) = \frac{2}{n}\sum_{i=1}^{n} x_{ij}\,\vec{x}_i^T\vec{w} - \frac{2}{n}\sum_{i=1}^{n} x_{ij}\,y_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_{ij}\,\vec{x}_i^T\vec{w} = \sum_{i=1}^{n} x_{ij}\,y_i$$

$$\sum_{i=1}^{n} x_{i0}\,\vec{x}_i^T\vec{w} = \sum_{i=1}^{n} x_{i0}\,y_i \qquad j = 0$$

$$\vdots \qquad\qquad \vdots$$

$$\sum_{i=1}^{n} x_{id}\,\vec{x}_i^T\vec{w} = \sum_{i=1}^{n} x_{id}\,y_i \qquad j = d$$

$$\underbrace{A^{-1}A}_{I}\,\vec{w} = A^{-1}\vec{b}$$

$$\hat{\vec{w}} = A^{-1}\vec{b}$$

$$A = \sum_{i=1}^{n} \underbrace{\vec{x}_i\vec{x}_i^T}_{\text{outer product}} \qquad \vec{b} = \sum_{i=1}^{n} \vec{x}_i\,y_i$$

Algorithm: Closed form Linear Regression Learner

input: $D = ((\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n))$

$$A = \sum_{i=1}^{n} \vec{x}_i \vec{x}_i^T$$

$$\vec{b} = \sum_{i=1}^{n} \vec{x}_i y_i$$

$$\hat{\vec{w}} = A^{-1} \vec{b}$$

return $\hat{f}(\vec{x}) = \vec{x}^T \hat{\vec{w}}$