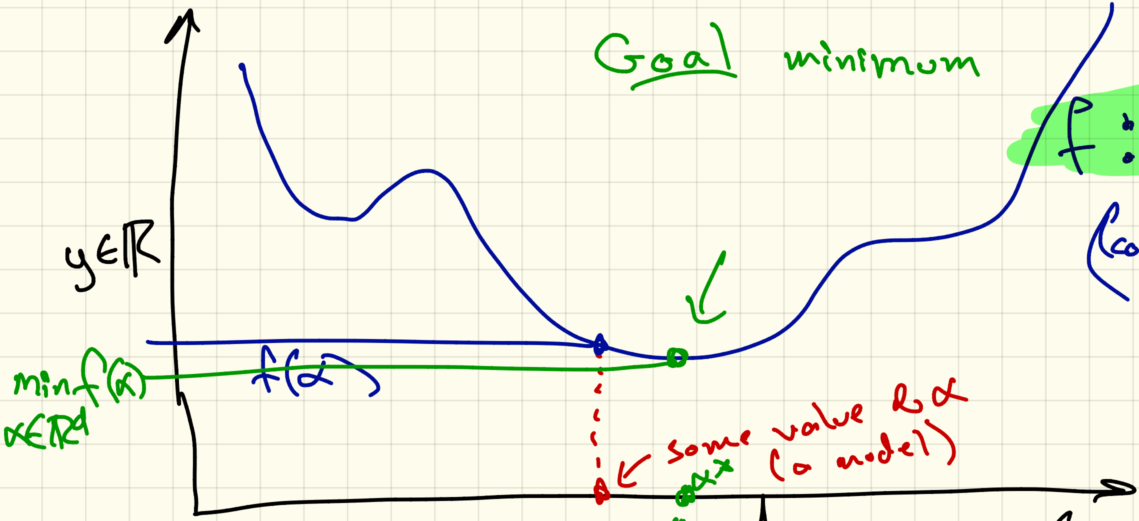
FoDA L15 : Gradient Descent (functions)

Goal: minimum



domain, range

$$f : \mathbb{R}^d \to \mathbb{R}$$

(cost function e.g. SSE)

$\alpha \in \mathbb{R}^d$ space of models

$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$

$$\alpha^* = \underset{\alpha \in \mathbb{R}^d}{\arg\min}\, f(\alpha)$$

example

$$f(\alpha) = SSE((x, y), M_\alpha) = \sum_{i=1}^{n} (y_i - \underbrace{\langle x_i, \alpha \rangle}_{M_\alpha(x_i)})^2$$

Properties & Defn of Functions

Local Neighborhood for $\alpha \in \mathbb{R}^d$

Euclidean ball $B_r(\alpha) = \{ p \in \mathbb{R}^d \mid \|\alpha - p\| \le r \}$

some ball $B_r(\alpha)$ for sufficiently small value $r > 0$.

---

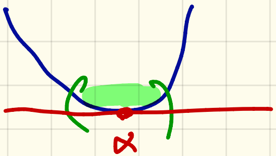
local minimum of f: a point $\alpha \in \mathbb{R}^d$

so for all $p \in B_r(\alpha)$

$$f(p) \ge f(\alpha)$$

"strict"

$$f(p) > f(\alpha), \quad p \neq \alpha$$



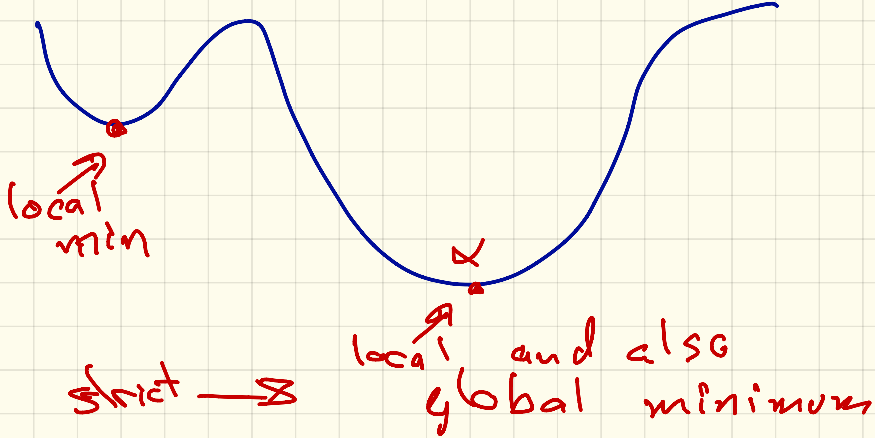
local maximum "

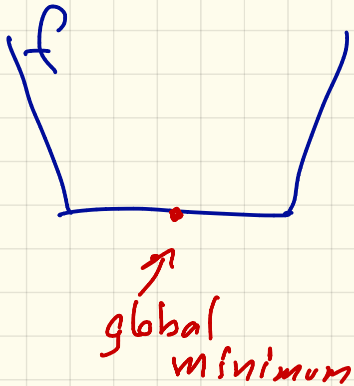
$$f(p) \le f(\alpha)$$

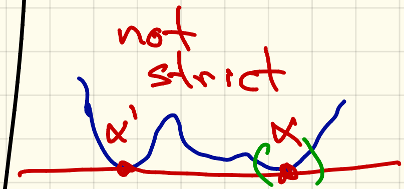
$$f(p) < f(\alpha), \quad p \neq \alpha$$

global minimum: $\alpha \in \mathbb{R}^d$ so $f(p) \geq f(\alpha)$ for all $p \in \mathbb{R}^d$

maximum $f(p) \leq f(\alpha)$ all $p \in \mathbb{R}^d$



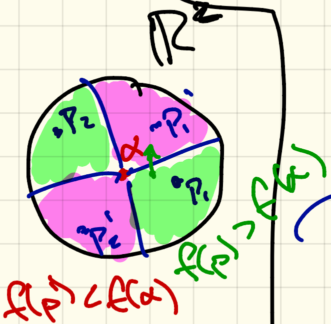




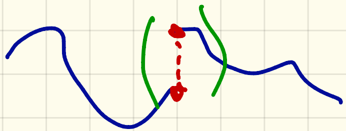
continuous function f if for any $\alpha \in \mathbb{R}^d$

$\exists$ sufficiently small $\delta > 0$ $\exists$ radius r_δ

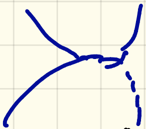
so $p \in B_{r_\delta}(\alpha)$ has $|f(\alpha) - f(p)| \leq \delta$



continuous



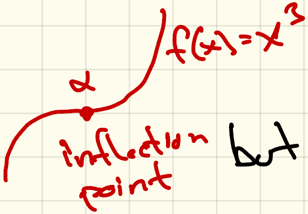
not continuous



saddle point $\alpha \in \mathbb{R}^d$ so in $B_r(\alpha)$

some p_1, p_2 w/ $f(p) > f(\alpha)$ some p_1', p_2' $f(p') < f(\alpha)$

but no path from p_1 to p_2 without passing through α or some p' s.t. $f(p') < f(\alpha)$

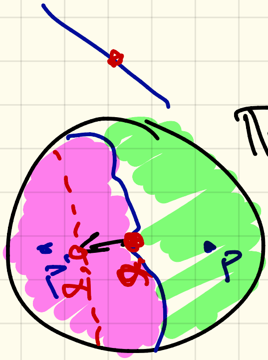


Most $\alpha \in \mathbb{R}^d$ are <u>regular points</u>

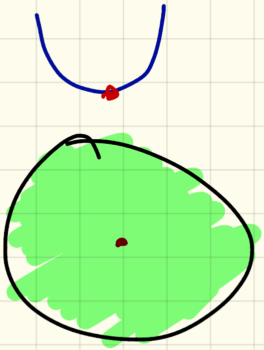
s.t. $\exists\, p, p' \in B_r(\alpha)$

so $f(p) > f(\alpha) > f(p')$

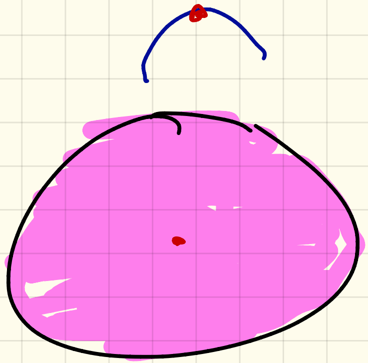
$\mathbb{R}^2$ and α <u>not</u> saddle point



regular point

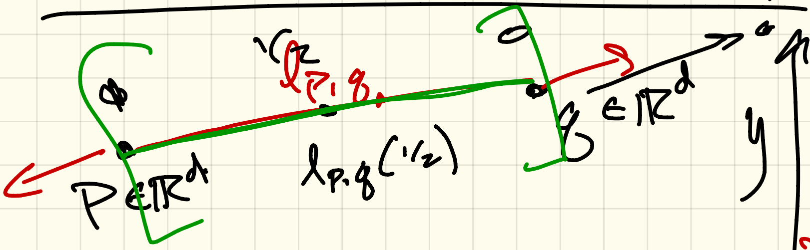


minimum

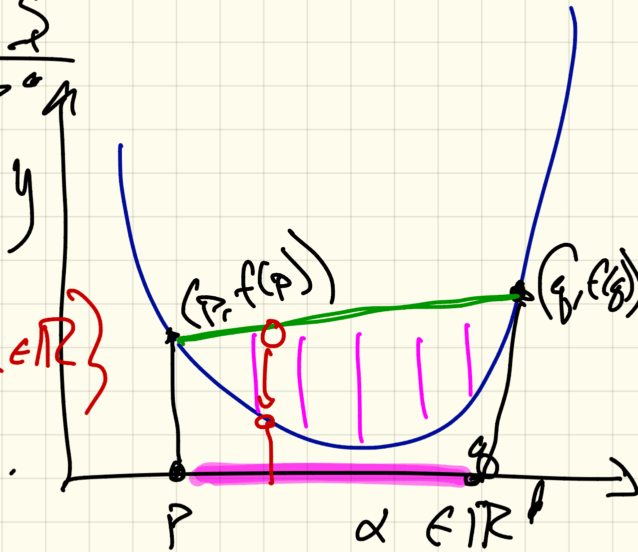


maximum

Convex functions



$$l_{p,q} = \{x = \lambda p + (1-\lambda) q \mid \lambda \in \mathbb{R}\}$$



$x = l_{p,q}(\lambda)$

$\lambda = 0$

$l_{p,q}(0) = q$

$l_{p,q}(1) = p$

$l_{p,q}(1/2) = \frac{p+q}{2}$

$l_{p,q}(-1)$

$\overline{l_{p,q}(\lambda)}$ for $\lambda \in [0,1]$

convex combination of p, q

$l_{p,q}(2)$

f is convex if $\overline{l_{(p,f(p)),(q,f(q))}}$ is "above" f

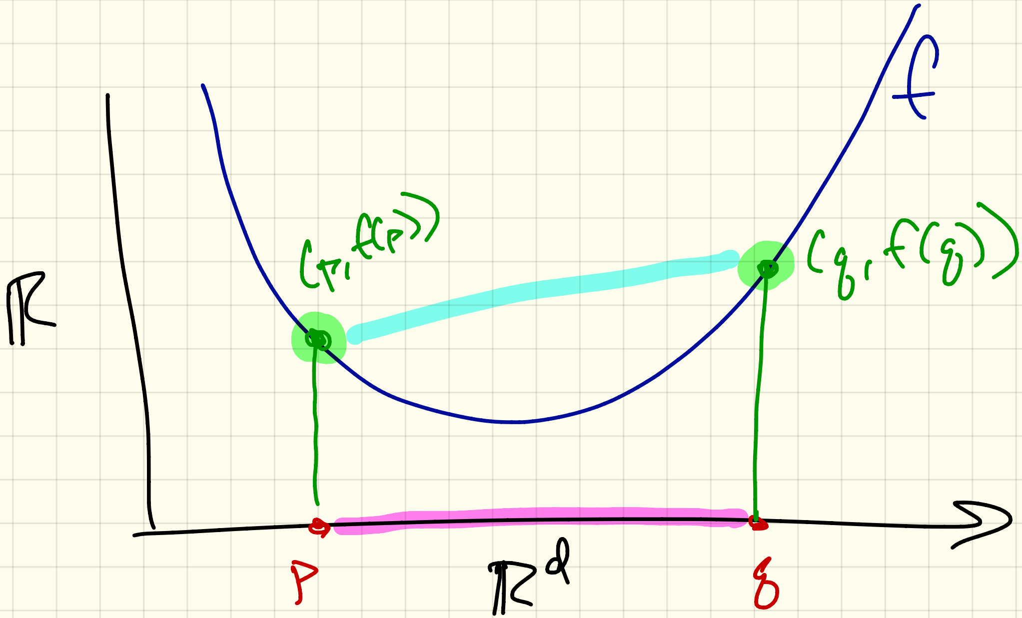
For all $p, q \in \mathbb{R}^d$ for all $\lambda \in [0,1]$

$$f(\lambda p + (1-\lambda) q) \le \lambda f(p) + (1-\lambda) f(q)$$

$f : \mathbb{R}^d \to \mathbb{R}$ is <u>convex</u> if

For all $p, q \in \mathbb{R}^d$ and all $\lambda \in [0,1]$

$$f(\lambda p + (1-\lambda) q) \leq \lambda f(p) + (1-\lambda) f(q)$$



Properties of Convex Functions

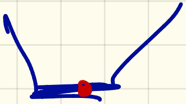
$$h(x) = f(x) + g(x)$$

and f, g convex $\Rightarrow$ h convex

$$h(x) = \max\{f(x), g(x)\}$$ then

f, g convex $\Rightarrow$ h convex

Any local minimum of convex f is



also a <u>global</u> minimum

Gradients

d dimension

∂ differential

if $\nabla_u f$ always defined

f differentiable

directional derivatives

$x \in (x_1 \ldots x_d)$

function $f : \mathbb{R}^d \to \mathbb{R}$

using "direction" $u \in \mathbb{R}^d$ $u = (u_1, u_2 \ldots u_d)$

$\|u\| = 1$

\$\nabla\$

$$\nabla_u f(x) = \lim_{h \to 0} \frac{f(x + hu) - f(x)}{h}$$

ith coord

$u = \{e_1, e_2 \ldots e_d\}$ $e_i = (0, 0, 0 \ldots, 0, 1, 0, \ldots 0)$

$$\nabla_i f(x) = \nabla_{e_i} f(x) = \frac{\partial}{\partial x_i} f(x)$$

gradient

$$\nabla f = \frac{\partial f}{\partial x_1} e_1 + \frac{\partial f}{\partial x_2} e_2 + \ldots = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_d}\right)$$

$\nabla f : \mathbb{R}^d \to \mathbb{R}^d$

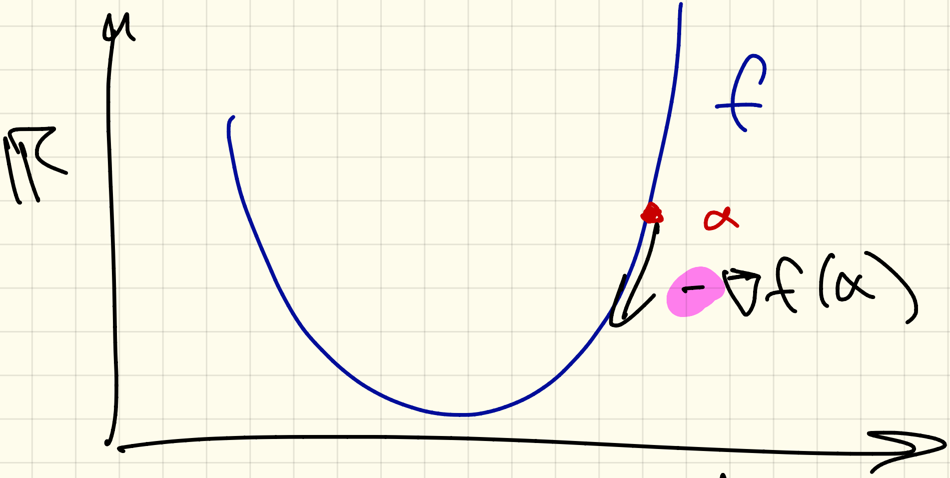
Example $\quad \alpha = (x, y, z) \in \mathbb{R}^3 \qquad e = 2.71\ldots$

$$f(\underset{\alpha_1}{x}, \underset{\alpha_2}{y}, \underset{\alpha_3}{z}) = 3x^2 - 2y^3 - 2xe^z$$

$$\frac{\partial f}{\partial x} = 6x - 2e^z \qquad \frac{\partial f}{\partial y} = -6y^2 \qquad \frac{\partial f}{\partial z} = -2xe^z$$

$$\nabla f = \left(6x - 2e^z, -6y^2, -2xe^z\right)$$

$$\nabla f(3, -2, 1) = \left(18 - 2e, -24, -6e\right)$$



steepest descent

$\alpha \in \mathbb{R}^d$

recover directional derivative

$$\nabla_v f(\alpha) = \langle \nabla f(\alpha), v \rangle$$

which direction v is $\nabla_v f(\alpha)$ largest?

$$\to \quad v = \frac{\nabla f(\alpha)}{\|\nabla f(\alpha)\|}$$