L14: Functions & Gradients

(and double descent in cross-validation)

Feb 25, 2026

FoDA

Jeff M. Phillips

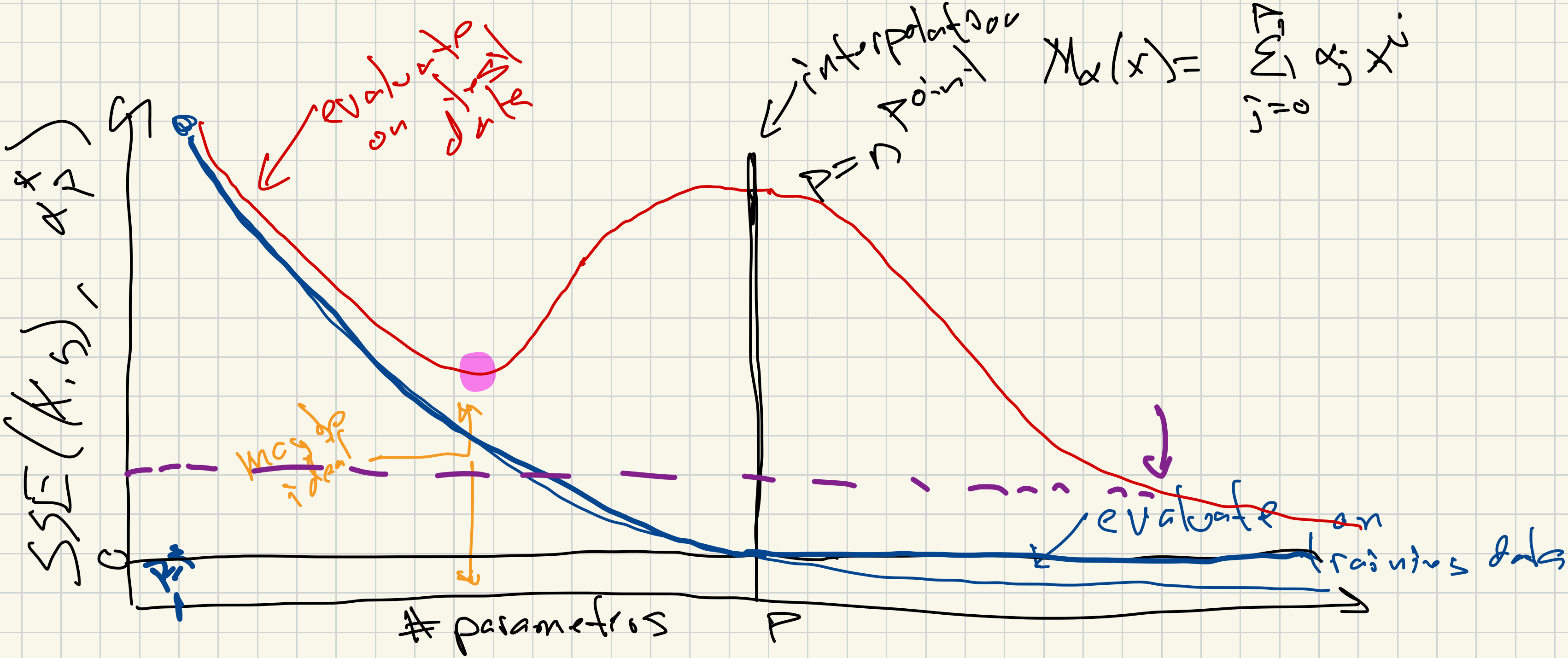
Cross Validation

polynomial regression

Input $(X, y) \to (X_T, y_T) \xrightarrow{\text{train}} \alpha^*_p = (\tilde{X}_p^T \tilde{X}_p)^{-1} \tilde{X}_p^T y$

$\xrightarrow{\text{random}} (X_E, y_E) \xrightarrow{\text{evaluate}} SSE((X_E, y_E), \alpha^*_p)$

$M_\alpha(x) = \sum_{j=0}^{p} \alpha_j x^j$



Functions

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

$f(\alpha) = SSE((X,y),\alpha)$

$\alpha = (\alpha_1, \alpha_2 \ldots \alpha_d) \in \mathbb{R}^d$

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

radius $r > 0$

local **Minimum** of f

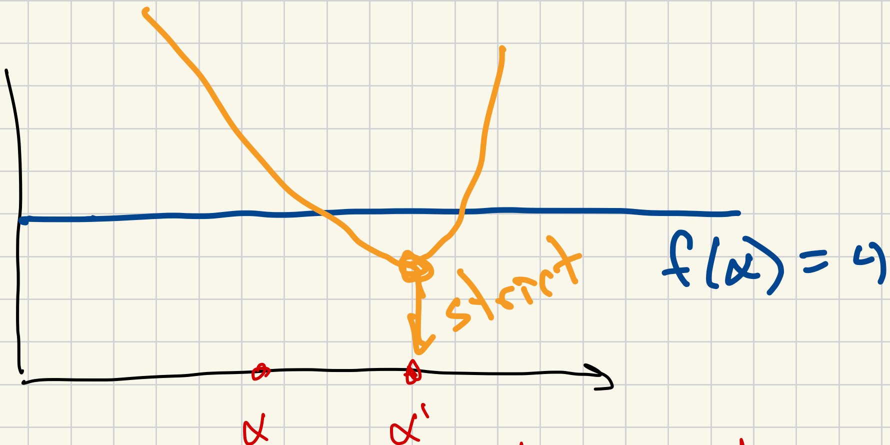
$\alpha \in \mathbb{R}^d$ s.t. all $p \in B_r(\alpha)$ $f(p) \ge f(\alpha)$

Global minimum of f: $\alpha \in \mathbb{R}^d$ s.t. all $p \in \mathbb{R}^d$ $f(p) \ge f(\alpha)$

maximum of f: $\alpha \in \mathbb{R}^d$ s.t.

$\forall p \in \cancel{B_r(\alpha)} \Rightarrow \mathbb{R}^d \Rightarrow$ global

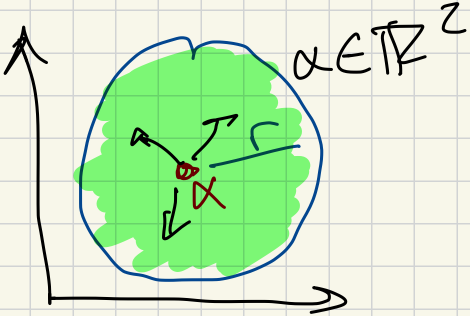
$f(p) \le f(\alpha)$ $<$ strict

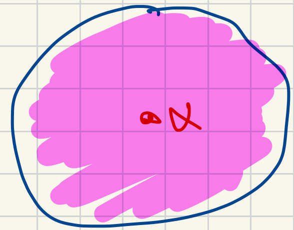


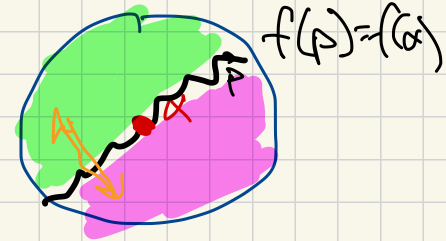
both global max
and global min

strict minimum of f: $\alpha \in \mathbb{R}^d$ s.t. $\forall p \in B_r(\alpha)$ $f(p) > f(\alpha)$

Draw in 2d



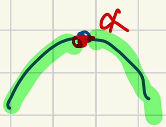


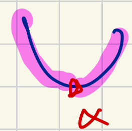


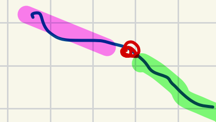
α maximum

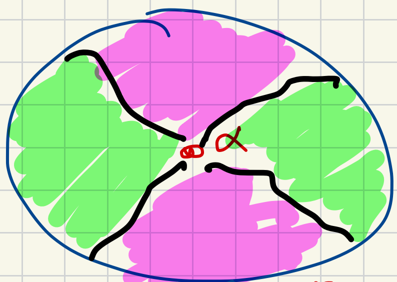
α minimum

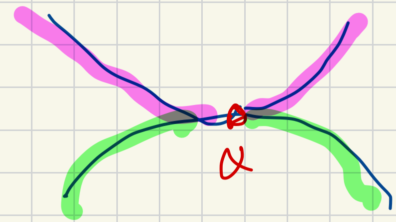
α regular







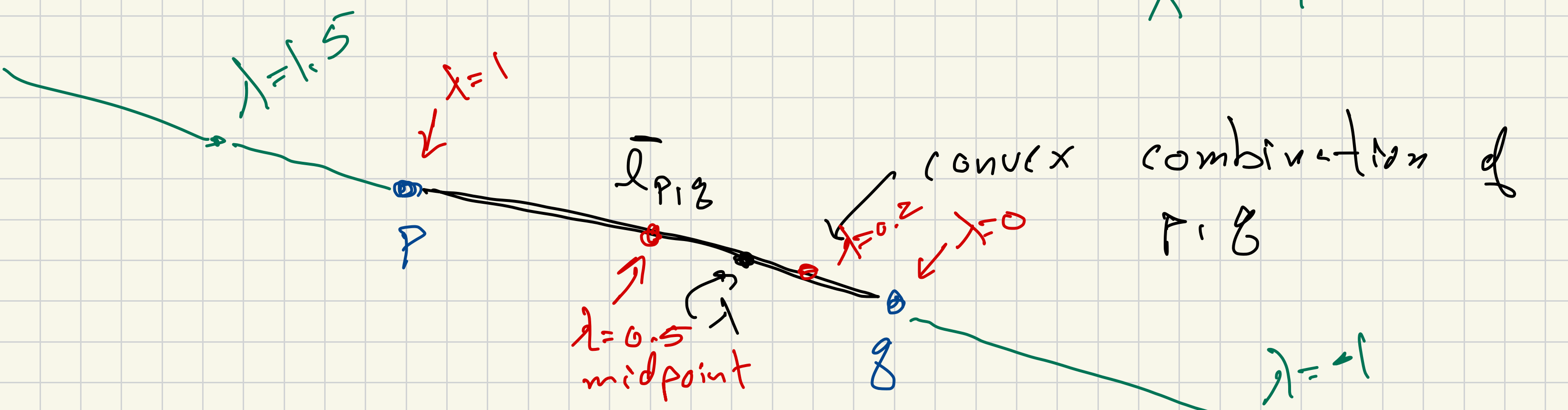




α = saddle point

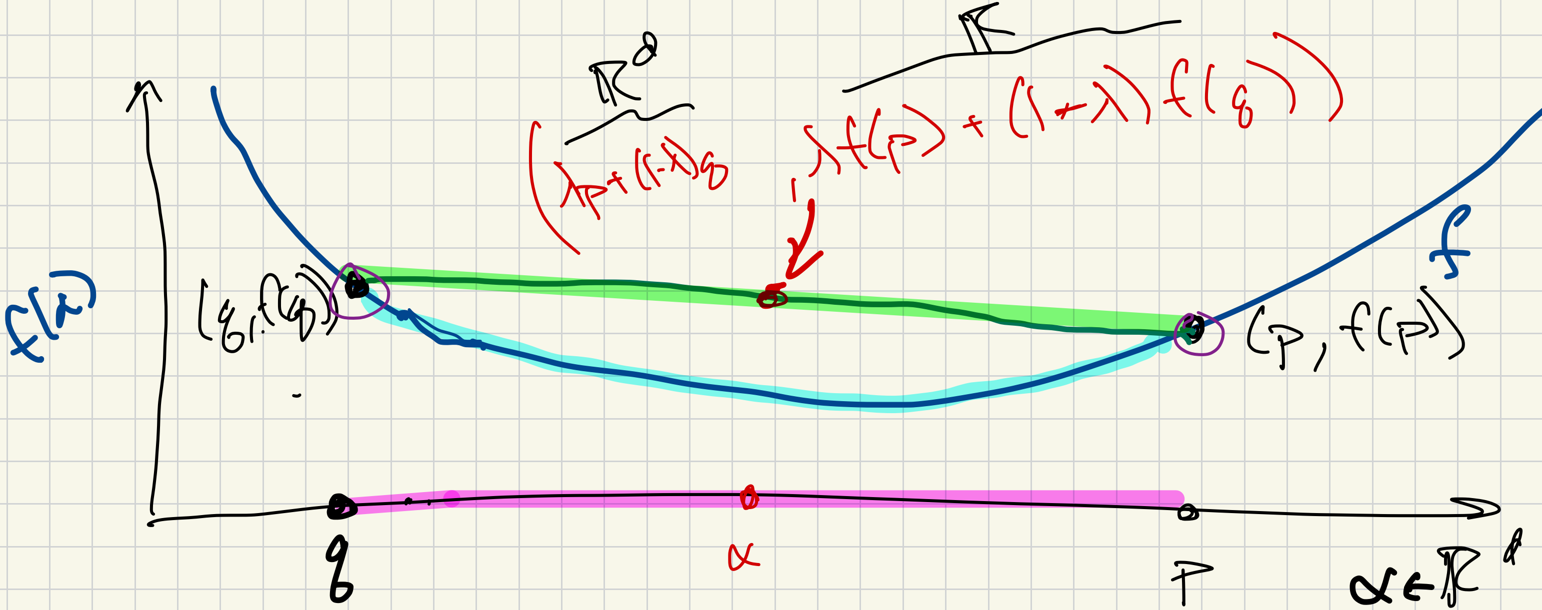
Convex Functions

$\lambda = -1$



$$\bar{l}_{p,q} = \{ x = \lambda p + (1-\lambda) q \mid \lambda \in [0,1] \}$$

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



f is <u>convex</u> if $\forall\ p, q \in \mathbb{R}^d$ all $\lambda \in [0,1]$

$\lambda \in (0,1)$

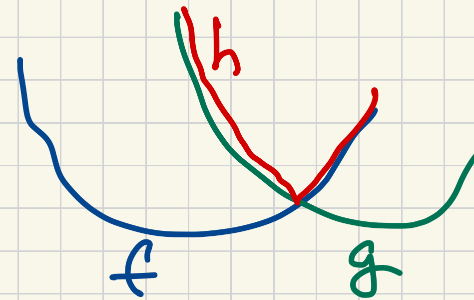
$$f(\underbrace{\lambda p + (1-\lambda) q}_{\alpha}) \le \lambda f(p) + (1-\lambda) f(q)$$

$<$ strict

Convex functions Properties

- f convex $\longrightarrow$ local minimum also global minimum

- composition
  f, g convex
  $\longrightarrow$ $h = f+g$ convex
  $\longrightarrow$ $h = \max\{f, g\}$ convex
  $\longrightarrow$ $h = f/\beta$ $\leftarrow$ const $\longrightarrow$ convex



Calculus → Derivatives

$$f(x) = f(x_1, x_2, \dots x_d)$$

unit vector

$v = (v_1, v_2 \dots v_d)$

$\|v\| = 1$

directional derivative

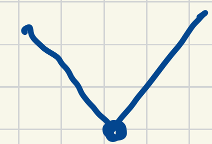
$$\nabla_v f(x) = \lim_{h \to 0} \frac{f(x + hv) - f(x)}{h}$$

\nabla

not continuous

if $\nabla_v f(x)$ is well-defined for all x, v

differentiably



unit vectors $e_1, e_2 \dots e_d$ ith coord.

$$e_i = (0, 0, 0, \dots 1, 0\ 0\ 0)$$

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

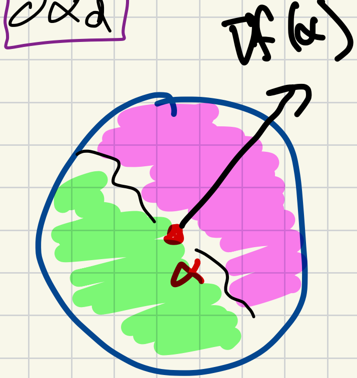
Gradient of f

scalar function

vector

$$\nabla f = \boxed{\frac{\partial f}{\partial x_1}}\,\boxed{e_1} + \boxed{\frac{\partial f}{\partial x_2}}\,\boxed{e_2} + \dots + \boxed{\frac{\partial f}{\partial x_d}}\,\boxed{e_d}$$

$$= \left(\boxed{\frac{\partial f}{\partial x_1}}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_d} \right)$$



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

$\Longrightarrow$ direction (+ scale) of steepest increase