Optimus

CS771: Introduction to Machine Learning

Purushottam Kar

Announcements

Assn groups: https://forms.gle/Zqe3yZyGv7rvzjm56 (deadline tonight)

Holiday: August 12 (Monday) is an institute holiday – no class

Quiz: August 14 (Wednesday), 6PM, L20

Note the new lecture hall for quiz (only for Aug 14 class)

Assigned seating – don't be late (will waste time finding your seat)

Syllabus is till whatever we cover today i.e. Aug 09 (Fri)

Bring your **institute ID card** with you – will lose time if you forget

Bring a **pencil, pen, eraser, sharpener** with you – we wont provide!

*Answers to be written on question paper itself. If you write with pen and make a mistake, no extra paper. Final answer **must be in pen***

Auditors cannot appear for quiz – please come to L20 at ~ 6:40PM

Class will resume after quiz is over (only 30 min quiz)

Recap of Last Lecture

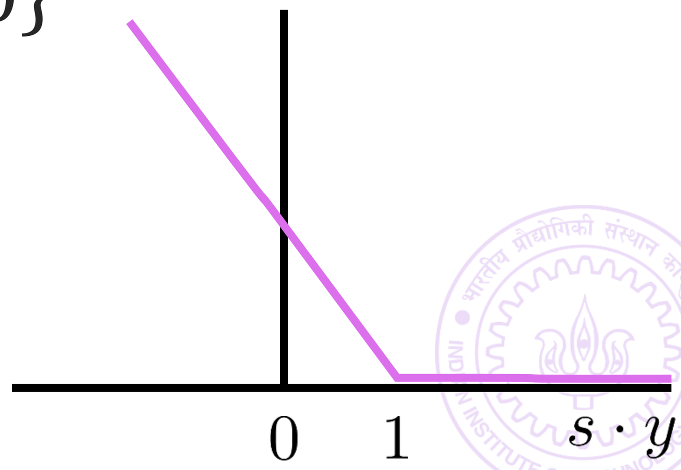
Looked at linear classifiers in more detail

Notions of *margin* – geometric margin, functional margin

Derived the SVM and C-SVM objectives as simply demanding a model that classify well but not let any data point come close to boundary

$$\min_{\tilde{\mathbf{w}},\tilde{b}} \frac{1}{2}\|\tilde{\mathbf{w}}\|_2^2 + C\sum_{i=1}^n \ell_{\text{hinge}}\left(\tilde{\mathbf{w}}^\top \mathbf{x}^i + \tilde{b}, y^i\right)$$

$$\ell_{\text{hinge}}(s,y) = \max\{1 - s\cdot y, 0\}$$



Recap of Last Lecture

Looked at linear classifiers in more detail

Notions of *margin* – geometric margin, functional margin

Derived the SVM and C-SVM objectives as simply demanding a model that classify well but not let any data point come close to boundary

$$\min_{\tilde{\mathbf{w}},\tilde{b}} \frac{1}{2}\|\tilde{\mathbf{w}}\|_2^2 + C\sum_{i=1}^{n} \ell_{\text{hinge}}\left(\tilde{\mathbf{w}}^\top \mathbf{x}^i + \tilde{b}, y^i\right)$$

$$\ell_{\text{hinge}}(s,y) = \max\{1 - s\cdot y, 0\}$$

Looked at what optimization problems look like – objective, constraints

$$\min_{x} x^2$$

s.t. $x \leq 6$ and $x \geq 3$

Recap of Last Lecture

Looked at linear classifiers in more detail

Notions of *margin* – geometric margin, functional margin

Derived the SVM and C-SVM objectives as simply demanding a model that classify well but not let any data point come close to boundary

$$\min_{\tilde{\mathbf{w}},\tilde{b}} \frac{1}{2}\|\tilde{\mathbf{w}}\|_2^2 + C\sum_{i=1}^{n} \ell_{\text{hinge}}\left(\tilde{\mathbf{w}}^\top \mathbf{x}^i + \tilde{b}, y^i\right)$$

$$\ell_{\text{hinge}}(s,y) = \max\{1 - s\cdot y, 0\}$$

Looked at what optimization problems look like – objective, constraints

$$\min_x x^2$$

$$\text{s.t. } x \leq 6 \text{ and } x \geq 3$$

Objective

Constraints

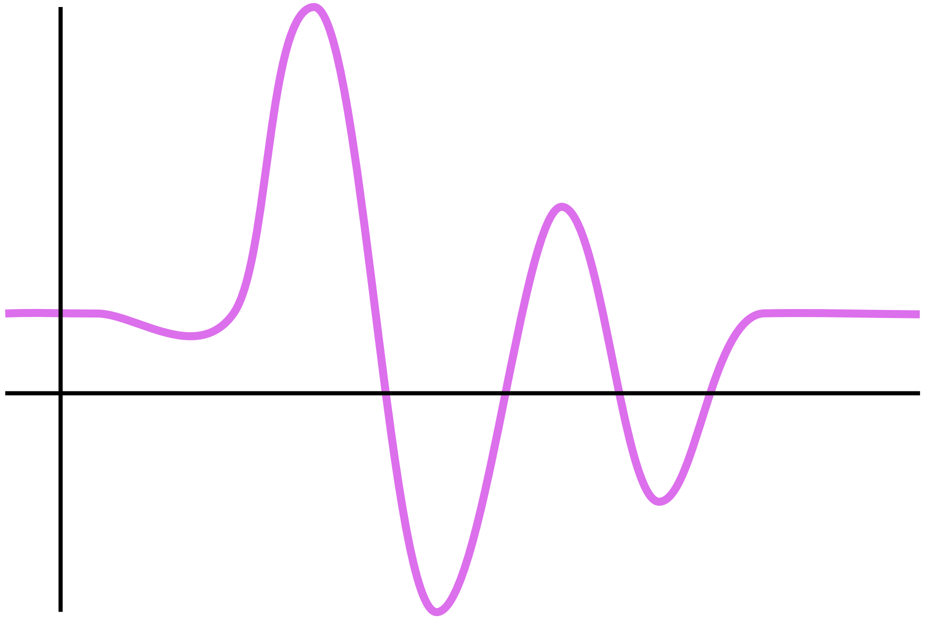
OPTIMization and calculUS

- Calculus basics and dealing with non-differentiable functions
- Convex sets and convex functions
- Gradient descent, sub-gradient descent, coordinate descent
- Lagrangian, dual optimization problems

Extrema

Since we always seek the "best" values of a function, usually we are looking for the maxima or the minima of a function

Global extrema: a point which achieves the best value of the function (max/min) among all the possible points



Local extrema: a point which achieves the best value of the function only in a small region surrounding that point

Most machine learning algorithms love to find the global extrema

- E.g. we saw that CSVM wanted to find the model with max margin

Sometimes it is difficult so we settle for local extrema (e.g. deepnets)

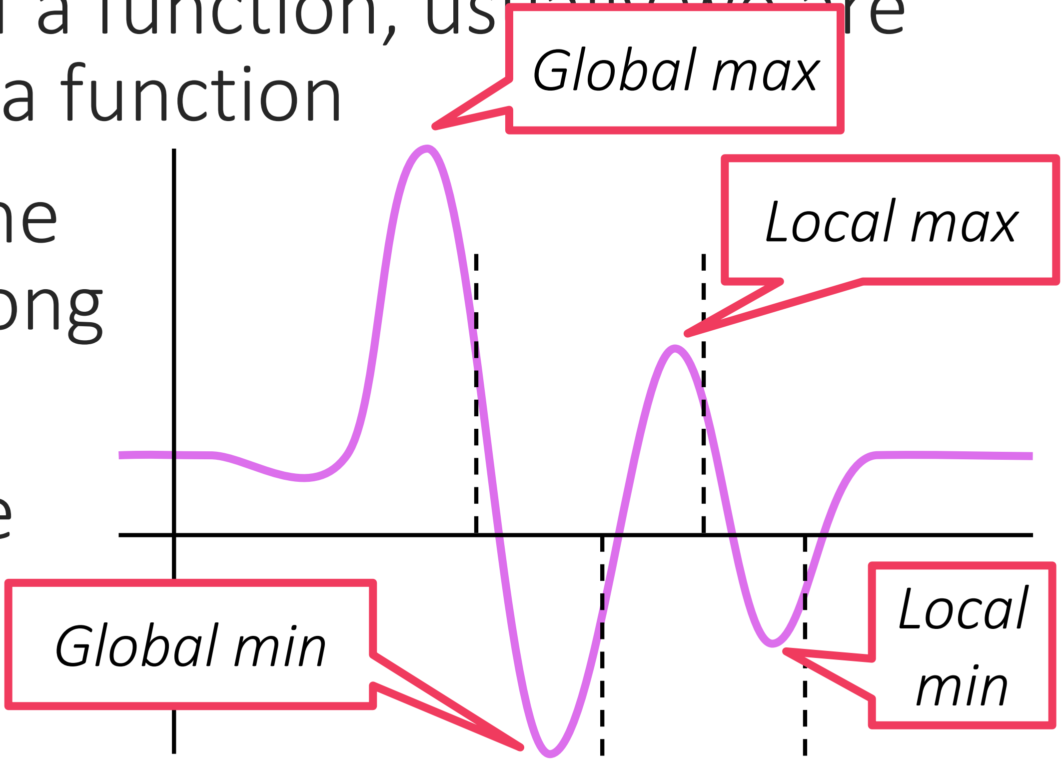
Extrema

Forget constraints for now – we will take care of them later!

Since we always seek the "best" values of a function, usually we are looking for the maxima or the minima of a function

Global extrema: a point which achieves the best value of the function (max/min) among all the possible points

Local extrema: a point which achieves the best value of the function only in a small region surrounding that point



Most machine learning algorithms love to find the global extrema

- E.g. we saw that CSVM wanted to find the model with max margin

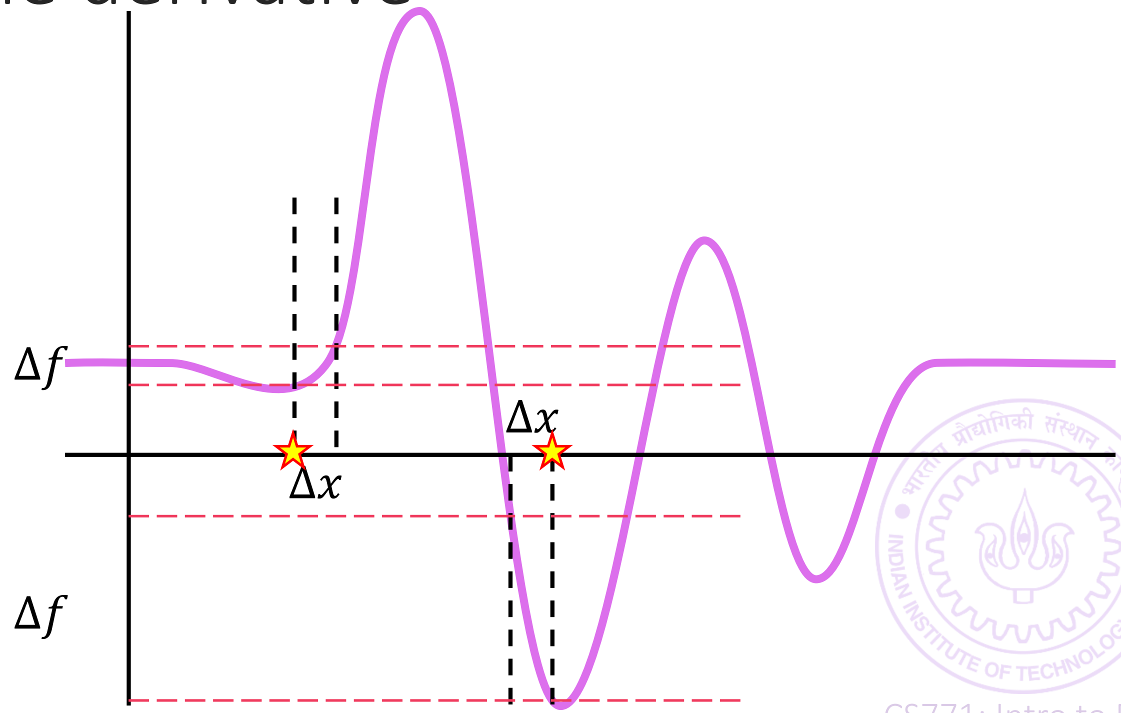
Sometimes it is difficult so we settle for local extrema (e.g. deepnets)

Derivatives

For a function $f: \mathbb{R} \to \mathbb{R}$, the sign of its derivative at any point tells us whether we should move left or right on the number line to *increase* f.

If sign is positive, we should move right else left

Magnitude of the derivative tells us how steeply would f increase if we moved a teeny tiny bit according to the derivative



Derivat

Derivatives only tell us how f will behave close to the point at which the derivative was calculated. If you move too much in direction of derivative, f may start decreasing. Similarly, if you move too much opposite to derivate, f may start increasing

oint tells us

r we should move left or right on the number line to *increase* f.

is positive,

Magnitude of the d … ld f increase if we moved a teeny tiny

Corollary of Taylor's Theorem

$$f(x + \Delta x) \approx f(x) + \Delta x \cdot f'(x)$$

if Δx is "small"

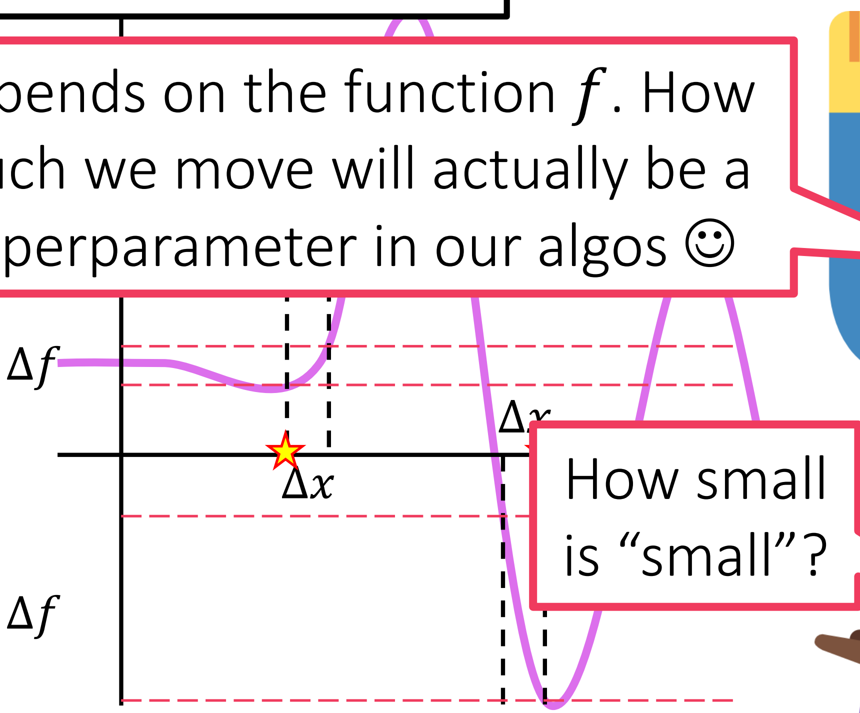
If we move a little bit opposite to the direction of derivative, then f would *decrease*

Depends on the function f. How much we move will actually be a hyperparameter in our algos ☺

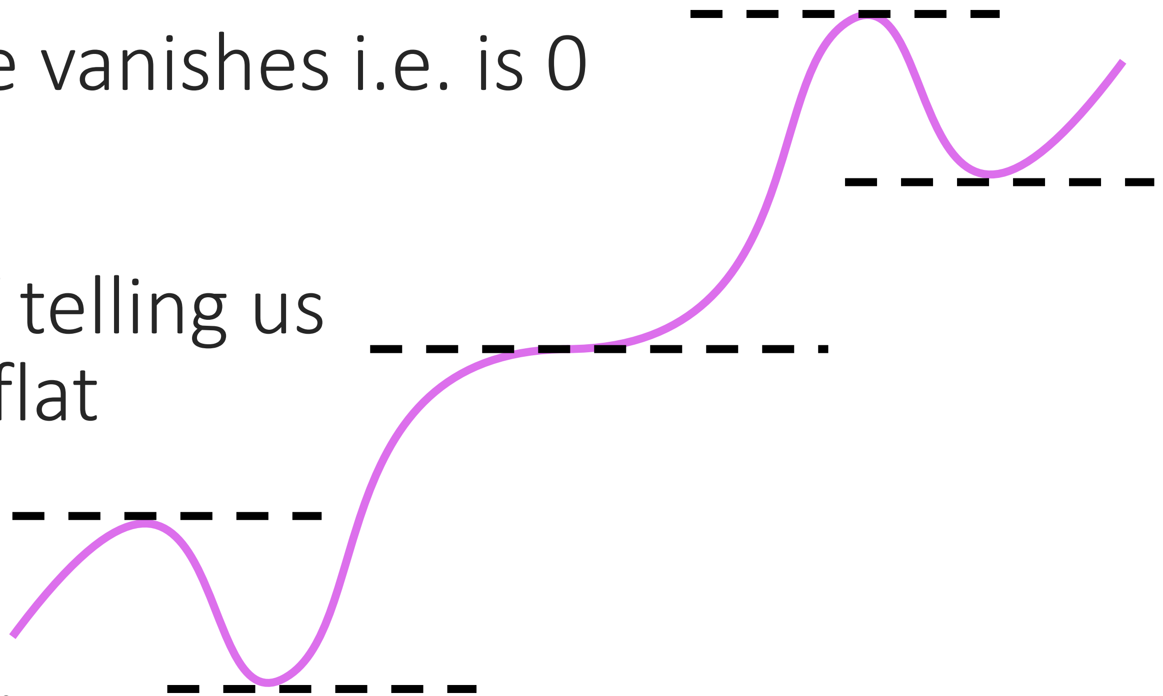
What if I moved in the opposite direction of the derivative?

Why do you keep saying "little bit"? What if I move a lot?

How small is "small"?



Stationary Points



These are places where the derivative vanishes i.e. is 0

These can be local/global extrema

The derivative being zero is its way of telling us that at that point, the function looks flat

Saddle points can be tedious in ML

We can find out if a stationary point is saddle or extrema using 2^{nd} derivative

Just as sign of the derivative tells us if the function is increasing or decreasing if we move left a tiny bit, the 2^{nd} derivative tells us if the derivative is increasing or decreasing if we move left a tiny bit

Stationary Points

If $f''(x) < 0$ and $f'(x) = 0$ then derivative moves from +ve to -ve around this point – local/global max!

These are places where the derivative vanishes i.e. is 0

If $f''(x) = 0$ and $f'(x) = 0$ then this may be extrema/saddle – higher derivatives e.g. $f'''(x)$ needed

These can be local

The derivative bei

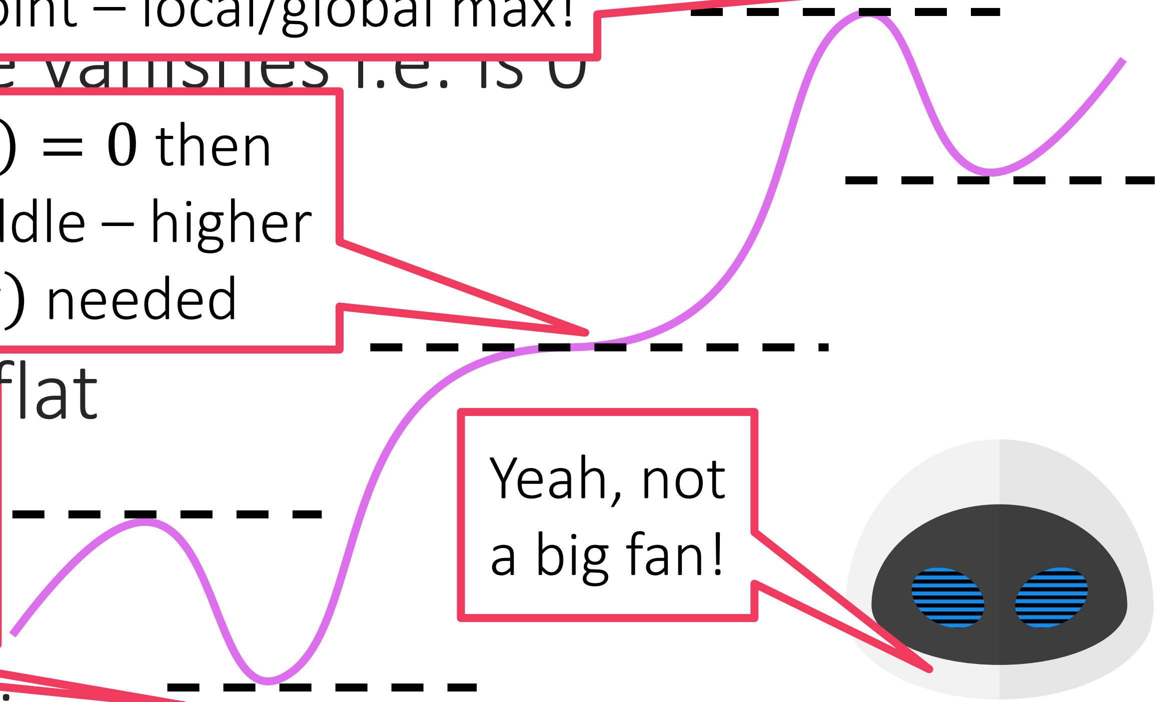
that at t flat

If $f''(x) > 0$ and $f'(x) = 0$ then derivative moves from -ve to +ve around this point – local/global min!

Saddle p

We can

is saddle or extrema using 2nd derivative



Yeah, not a big fan!

Just as sign of the derivative tells us if the function is increasing or decreasing if we move left a tiny bit, the 2nd derivative tells us if the derivative is increasing or decreasing if we move left a tiny bit

Rules of derivatives

Sum Rule: $\left(f(x)+g(x)\right)' = f'(x)+g'(x)$

Scaling Rule: $\left(a\cdot f(x)\right)' = a\cdot f'(x)$ if a is not a function of x

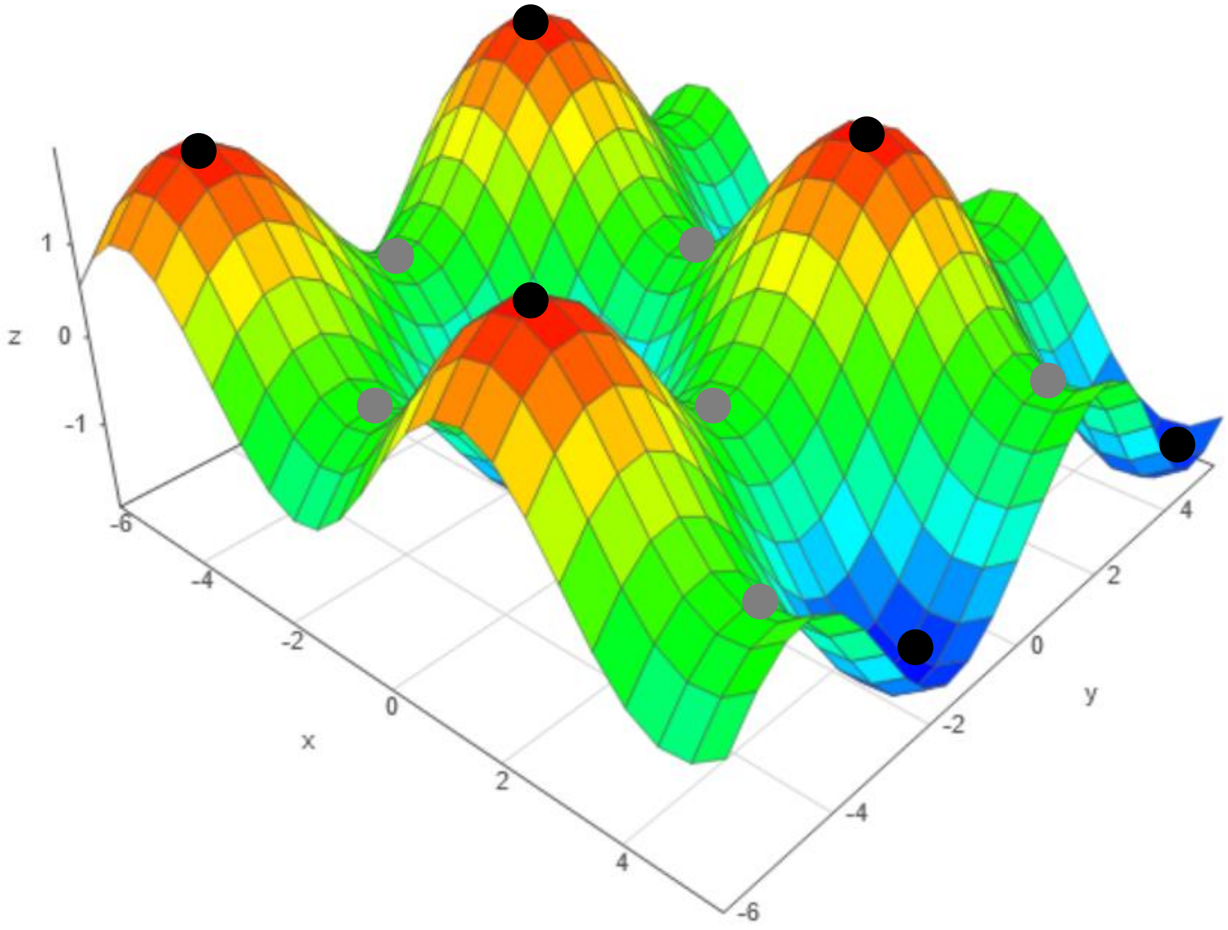
Product Rule: $\left(f(x)\cdot g(x)\right)' = f'(x)\cdot g(x)+g'(x)\cdot f(x)$

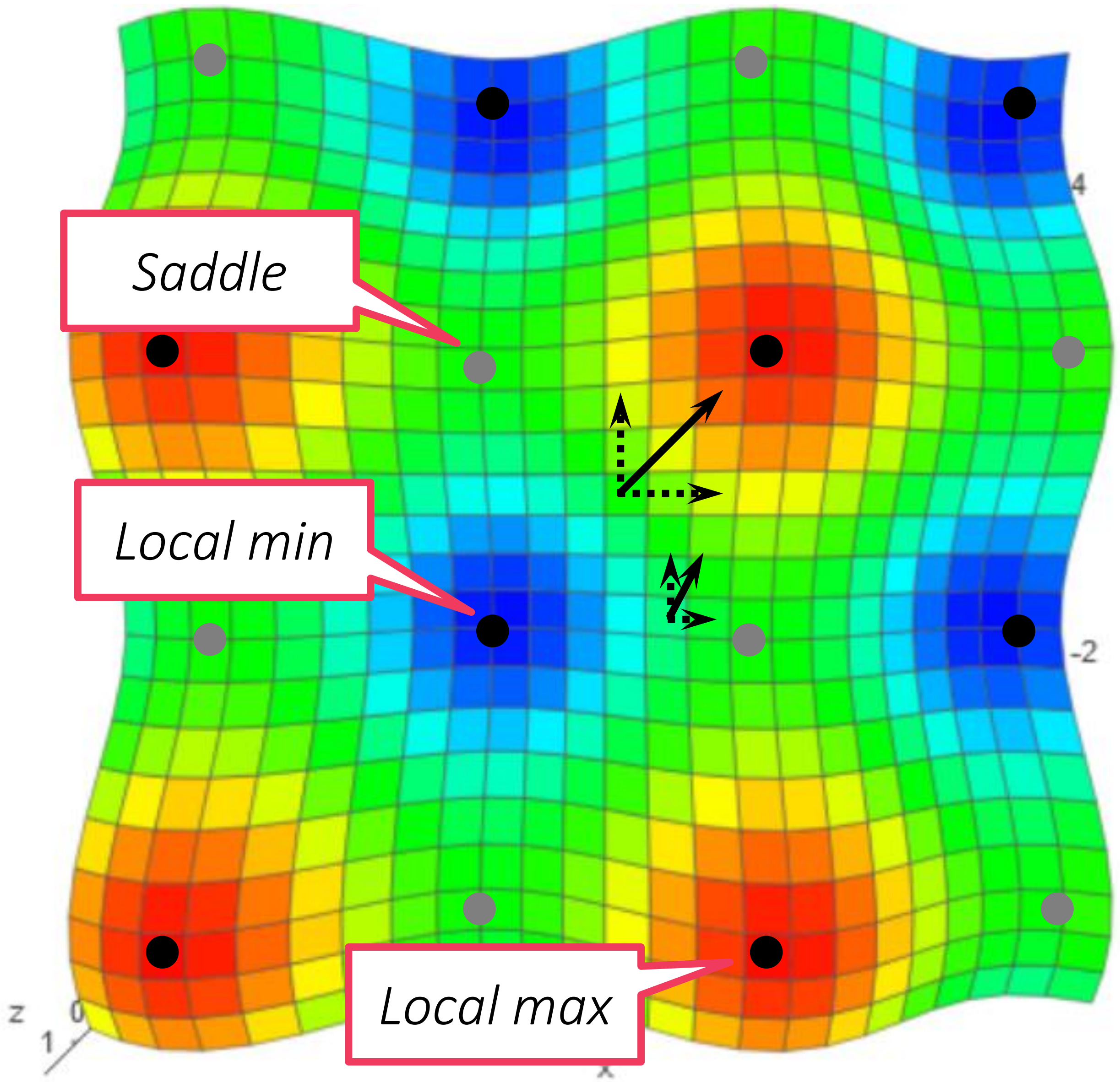
Quotient Rule: $\left(f(x)/g(x)\right)' = \left(f'(x)\cdot g(x)-g'(x)f(x)\right)/\left(g(x)\right)^2$

Chain Rule: $\left(f\left(g(x)\right)\right)' \overset{\text{def}}{=} (f\circ g)'(x) = f'\left(g(x)\right)\cdot g'(x)$

Most common use f is a function of t but $t = g(x)$, calculate df/dx

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





PLOTS COURTESY ACADEMO.ORG

Multivariate Function

This looks just like the 1D case except that we are summing up contributions from all d dimensions

Gradient

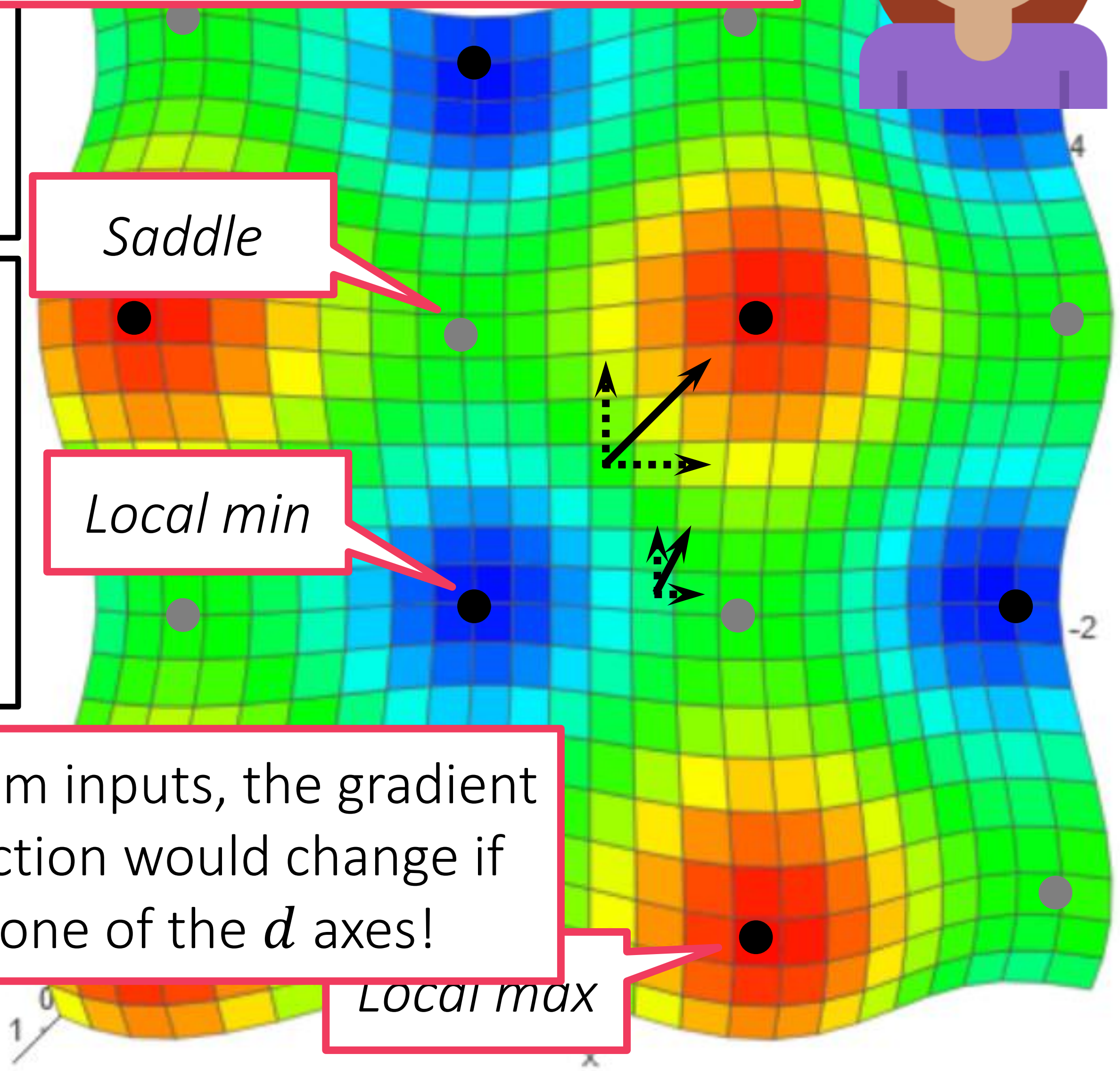
$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_d}\right)$$

Taylor's Theorem in higher dims

If we move along vector $\mathbf{t} = (t_1, t_2, \ldots, t_d)$

then $f(\mathbf{x}+\mathbf{t}) \approx f(\mathbf{x}) + \sum_{i=1}^{d} t_i \cdot \frac{\partial f(\mathbf{x})}{\partial x_i} =$

$f(\mathbf{x}) + \mathbf{t}^\top \nabla f(\mathbf{x})$ if $\mathbf{t}$ is "small"



For multivariate functions with d-dim inputs, the gradient simply records how much the function would change if we move a little bit along each one of the d axes!

PLOTS COURTESY ACADEMO.ORG

Multivariate Functio

This looks just like the 1D case except that we are summing up contributions from all d dimensions

Gradient

∇f

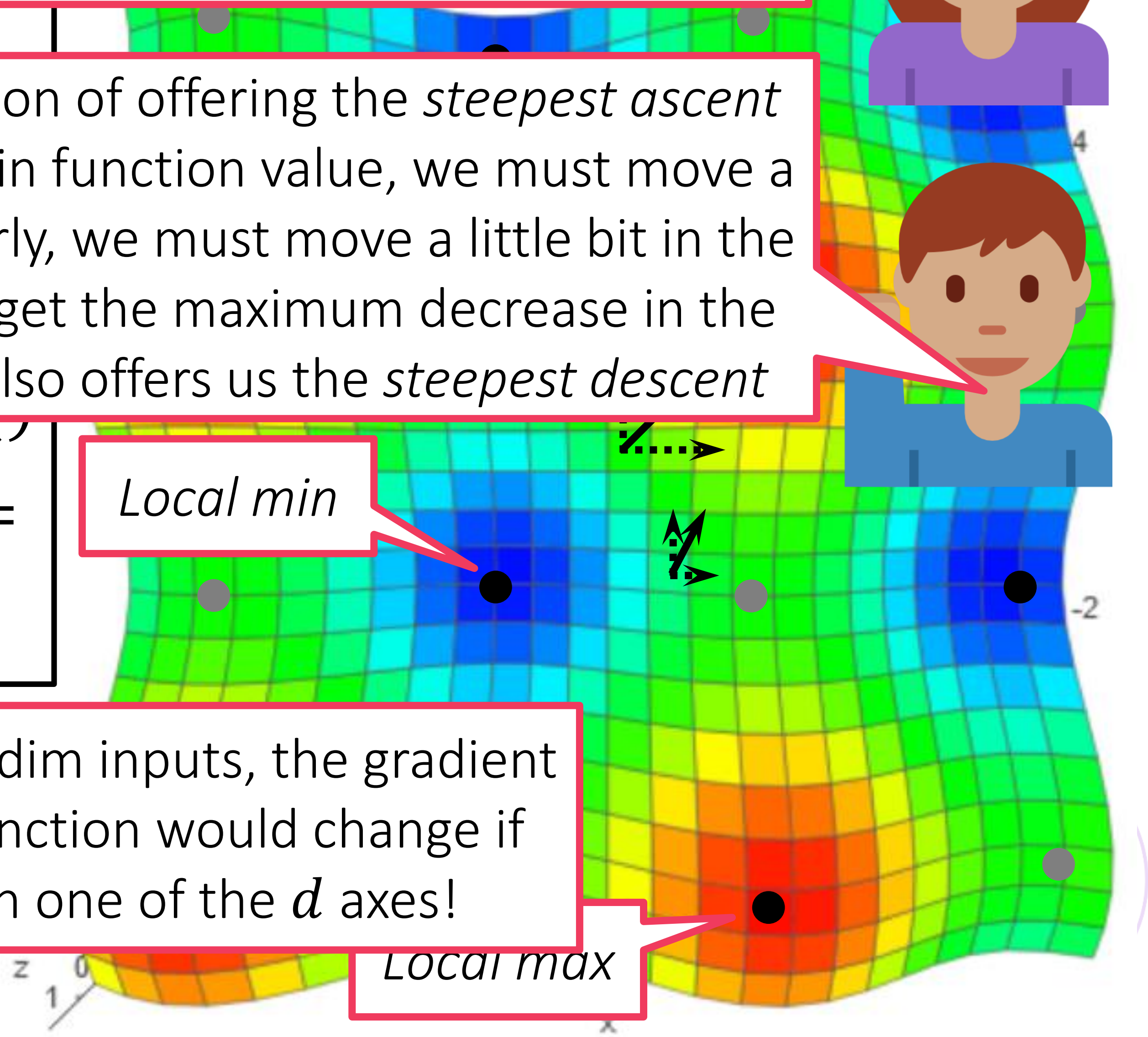
The gradient also has the distinction of offering the *steepest ascent* i.e. if we want maximum increase in function value, we must move a little bit along the gradient. Similarly, we must move a little bit in the direction opposite to gradient to get the maximum decrease in the function value, i.e. the gradient also offers us the *steepest descent*

Taylor's

If we move

then $f(\mathbf{x}+\mathbf{t}) \approx f(\mathbf{x}) + \sum_{i=1}^{d} t_i \cdot \frac{\partial f(\mathbf{x})}{\partial x_i} =$

$f(\mathbf{x}) + \mathbf{t}^\top \nabla f(\mathbf{x})$ if $\mathbf{t}$ is "small"



For multivariate functions with d-dim inputs, the gradient simply records how much the function would change if we move a little bit along each one of the d axes!

PLOTS COURTESY ACADEMO.ORG

Higher derivatives in higher dimensions

2nd derivative of $f: \mathbb{R}^d \to \mathbb{R}$ is a $d \times d$ matrix called the *Hessian*

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_d} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d x_1} & \frac{\partial^2 f}{\partial x_d x_2} & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}$$

May get difficult to *visualize* higher derivatives – just go with the math

3rd and higher derivatives must be expressed as *tensors*

All rules of derivatives (chain, product etc) apply here as well

Stationary Points in d-dimensions

These are places where the gradient vanishes i.e. is a zero vector!

We can still find out if a stationary point is saddle or extrema using the 2nd derivative test just as in 1D

A bit more complicated to visualize, but the Hessian tells us how the surface of the function is curved at a point

If $\nabla f(\mathbf{x}) = \mathbf{0}$ and $\nabla^2 f(\mathbf{x})$ is a PSD matrix, then $\mathbf{x}$ is a local/global min

If $\nabla f(\mathbf{x}) = \mathbf{0}$ and $\nabla^2 f(\mathbf{x})$ is a NSD matrix, then $\mathbf{x}$ is a local/global max

Else test fails, need higher order derivatives to verify

Stationary Points in d-dimensions

These are places where the gradient vanishes i.e. is a zero vector!

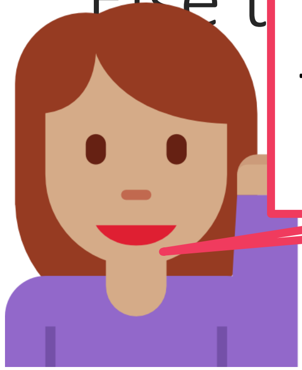
We can still find out if a stationary point is saddle or extrema using the 2nd derivative test just as in 1D

A bit more complicated to visualize, but the Hessian tells us how the surface of the function is curved at a point

If $\nabla f(\mathbf{x}) = \mathbf{0}$ and $\nabla^2 f(\mathbf{x})$ is a PSD matrix, then $\mathbf{x}$ is a local/global min

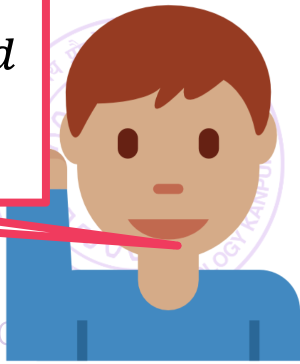
If $\nabla f(\mathbf{x}) = \mathbf{0}$ and $\nabla^2 f(\mathbf{x})$ is a NSD matrix, then $\mathbf{x}$ is a local/global max

Else t ... de



If a matrix satisfies $\mathbf{x}^\top A\mathbf{x} \leq 0$ for all $\mathbf{x} \in \mathbb{R}^d$ then it is called *negative semidefinite (NSD)*

Recall that if a square $d \times d$ symmetric matrix A satisfies $\mathbf{x}^\top A\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^d$ then it is *positive semidefinite (PSD)*



A Toy Example – Function Values

6	3	3	3	3	3	3	3	3	3
5	2	2	2	3	4	3	3	2	1
4	1	1	1	3	3	3	1	1	1
3	1	0	1	1	2	1	1	0	1
2	1	1	1	3	3	3	1	1	1
1	1	2	3	3	4	3	2	2	2
0	3	3	3	3	3	3	3	3	3
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

In this discrete toy example, we can calculate gradient at a point (x_0, y_0) as

$$\nabla f(x_0, y_0) = \left(\frac{\Delta f}{\Delta x}, \frac{\Delta f}{\Delta y}\right) \text{ where}$$

$$\frac{\Delta f}{\Delta x} = \frac{f(x_0+1, y_0) - f(x_0-1, y_0)}{2}$$

$$\frac{\Delta f}{\Delta y} = \frac{f(x_0, y_0+1) - f(x_0, y_0-1)}{2}$$

A Toy Example – Gradients

	0	1	2	3	4	5	6	7	8
6									
5		$(0,1)$	$\left(\frac{1}{2},1\right)$	$(1,0)$	$(0,0)$	$\left(-\frac{1}{2},0\right)$	$\left[-\frac{1}{2},1\right]$	$(-1,1)$	
4		$(0,1)$	$\left(1,\frac{1}{2}\right)$	$(1,1)$	$(0,1)$	$(-1,1)$	$(-1,1)$	$(0,1)$	
3		$(0,0)$	$\left(\frac{1}{2},0\right)$	$\left(\frac{1}{2},0\right)$	$(0,0)$	$\left(-\frac{1}{2},0\right)$	$\left(-\frac{1}{2},0\right)$	$(0,0)$	
2		$(0,-1)$	$(1,-1)$	$(1,-1)$	$(0,-1)$	$(-1,-1)$	$\left[-1,-\frac{1}{2}\right]$	$(0,-1)$	
1		$(1,-1)$	$\left(\frac{1}{2},-1\right)$	$\left(\frac{1}{2},0\right)$	$(0,0)$	$(-1,0)$	$\left[-\frac{1}{2},-1\right]$	$(0,-1)$	
0									

In this discrete toy example, we can calculate gradient at a point (x_0, y_0) as

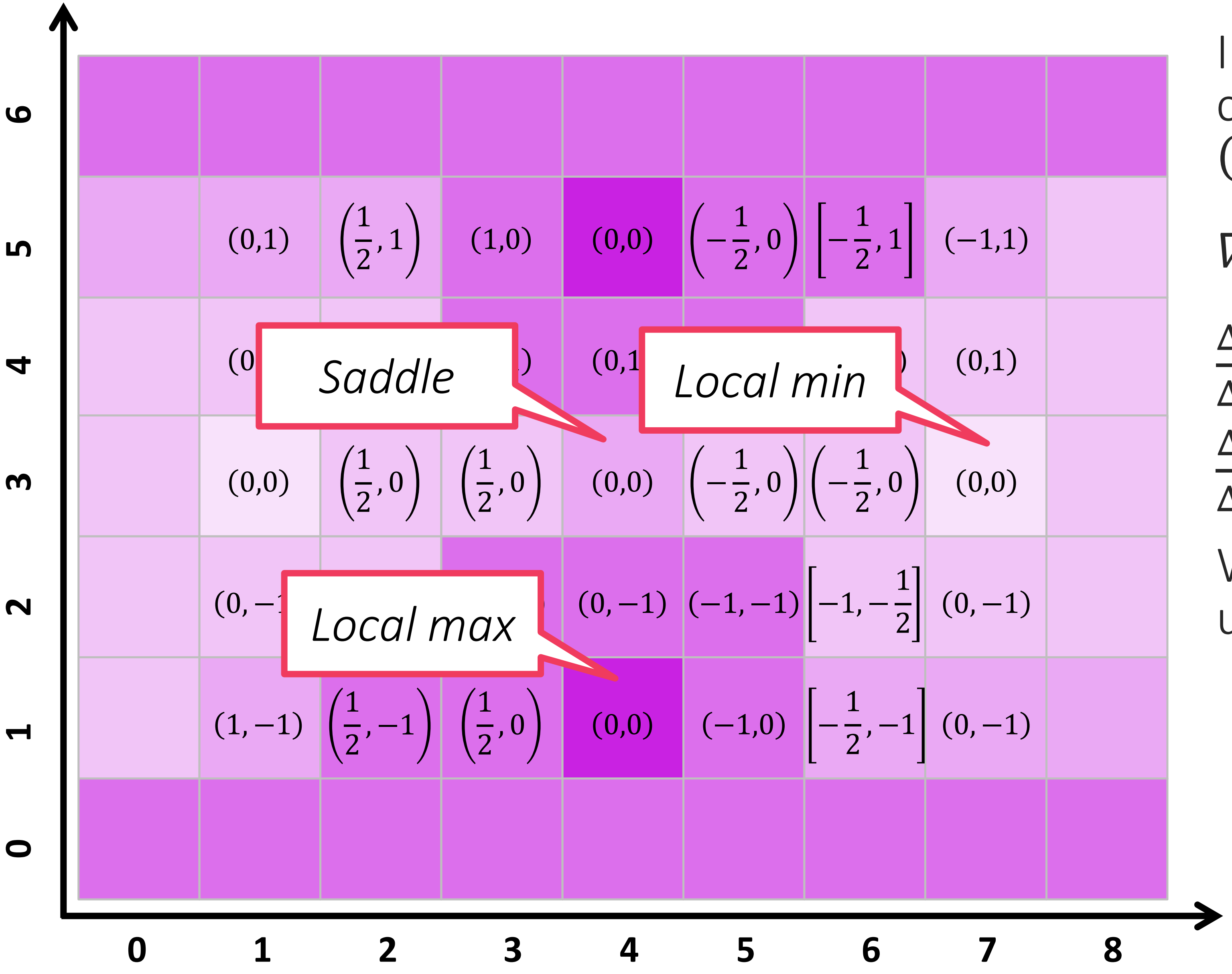
$$\nabla f(x_0, y_0) = \left(\frac{\Delta f}{\Delta x}, \frac{\Delta f}{\Delta y}\right) \text{ where}$$

$$\frac{\Delta f}{\Delta x} = \frac{f(x_0+1, y_0) - f(x_0-1, y_0)}{2}$$

$$\frac{\Delta f}{\Delta y} = \frac{f(x_0, y_0+1) - f(x_0, y_0-1)}{2}$$

We can visualize these gradients using simple arrows as well

A Toy Example – Gradients



In this discrete toy example, we can calculate gradient at a point (x_0, y_0) as

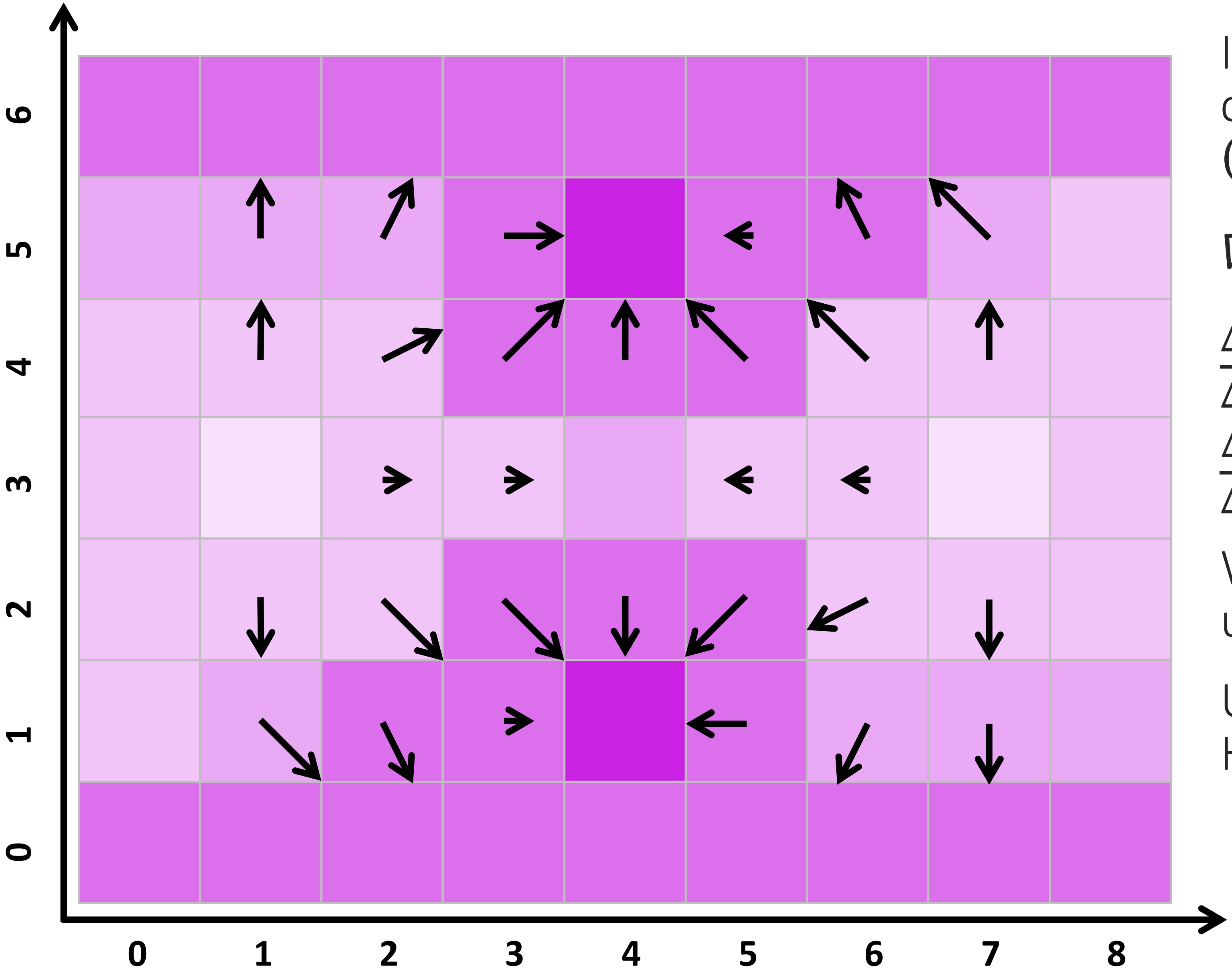
$$\nabla f(x_0, y_0) = \left(\frac{\Delta f}{\Delta x}, \frac{\Delta f}{\Delta y}\right) \text{ where}$$

$$\frac{\Delta f}{\Delta x} = \frac{f(x_0+1, y_0) - f(x_0-1, y_0)}{2}$$

$$\frac{\Delta f}{\Delta y} = \frac{f(x_0, y_0+1) - f(x_0, y_0-1)}{2}$$

We can visualize these gradients using simple arrows as well

A Toy Example – Gradients



In this discrete toy example, we can calculate gradient at a point (x_0, y_0) as

$$\nabla f(x_0, y_0) = \left(\frac{\Delta f}{\Delta x}, \frac{\Delta f}{\Delta y}\right) \text{ where}$$

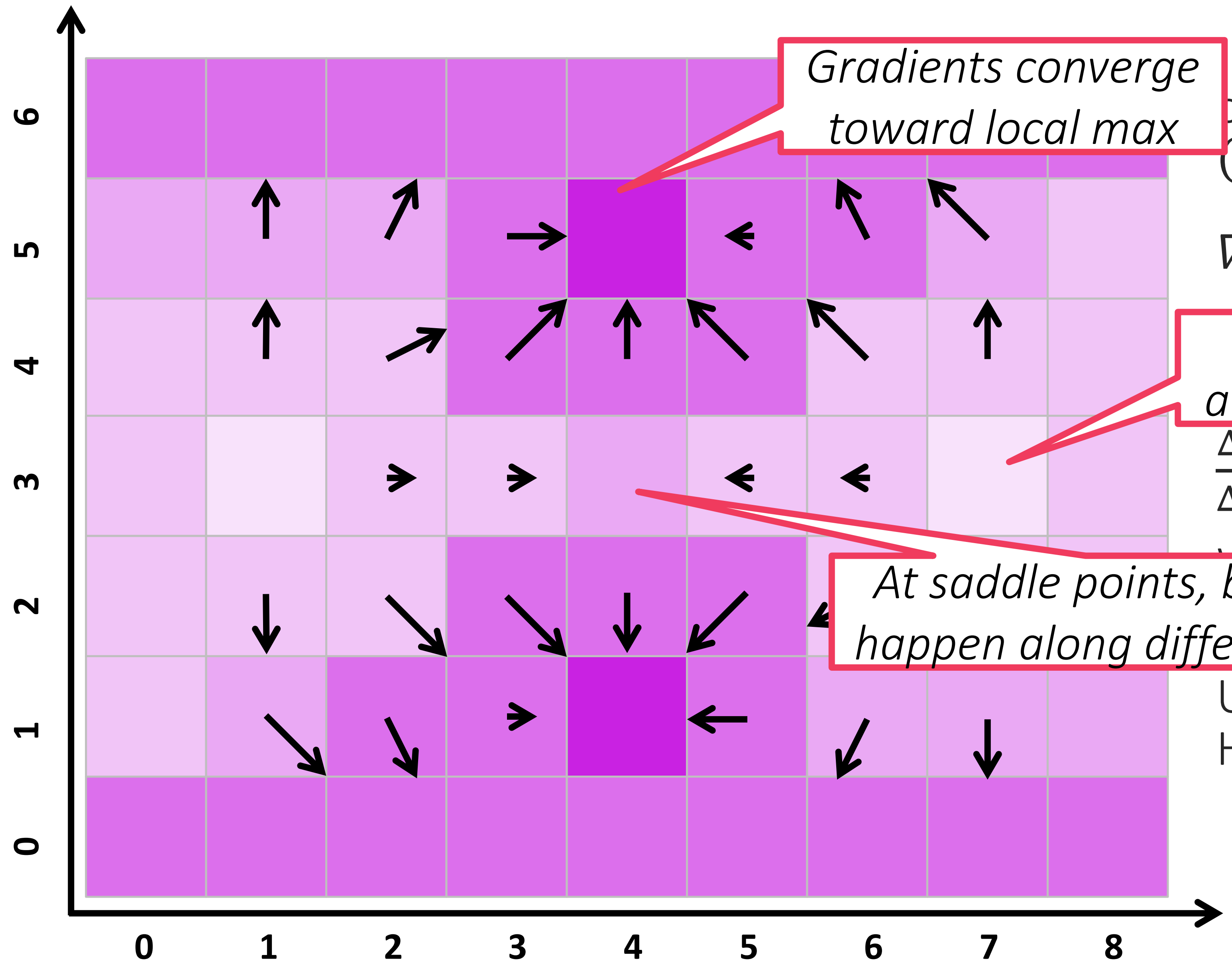
$$\frac{\Delta f}{\Delta x} = \frac{f(x_0+1, y_0) - f(x_0-1, y_0)}{2}$$

$$\frac{\Delta f}{\Delta y} = \frac{f(x_0, y_0+1) - f(x_0, y_0-1)}{2}$$

We can visualize these gradients using simple arrows as well

Using a similar method, the Hessian can be calculated as well!

A Toy Example – Gradients



n this discrete toy example, we can calculate gradient at a point (x_0, y_0) as

$\nabla f(x_0, y_0) = \left(\frac{\Delta f}{\Delta x}, \frac{\Delta f}{\Delta y}\right)$ where

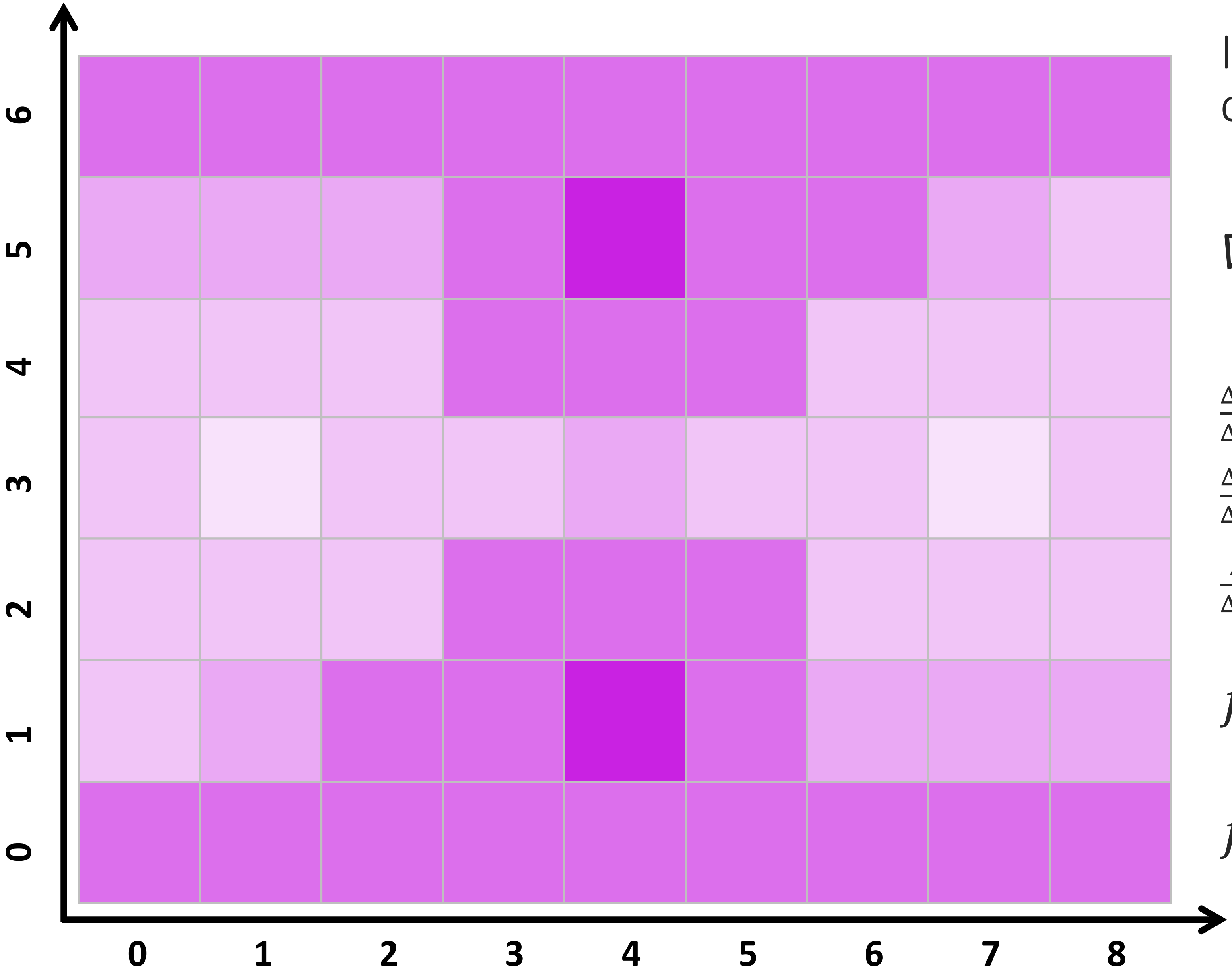
$-1, y_0)$

$$\frac{\Delta f}{\Delta y} = \frac{f(x_0, y_0+1) - f(x_0, y_0-1)}{2}$$

ualize these gradients e arrows as well

Using a similar method, the Hessian can be calculated as well!

A Toy Example – Hessians



In this discrete toy example, we can calculate Hessian at (x_0, y_0) as

$$\nabla^2 f(x_0, y_0) = \begin{bmatrix} \frac{\Delta^2 f}{\Delta x^2} & \frac{\Delta^2 f}{\Delta x \Delta y} \\ \frac{\Delta^2 f}{\Delta x \Delta y} & \frac{\Delta^2 f}{\Delta y^2} \end{bmatrix} \text{ where}$$

$$\frac{\Delta^2 f}{\Delta x^2} = f(x_0 + 1, y_0) + f(x_0 - 1, y_0) - 2f(x_0, y_0)$$

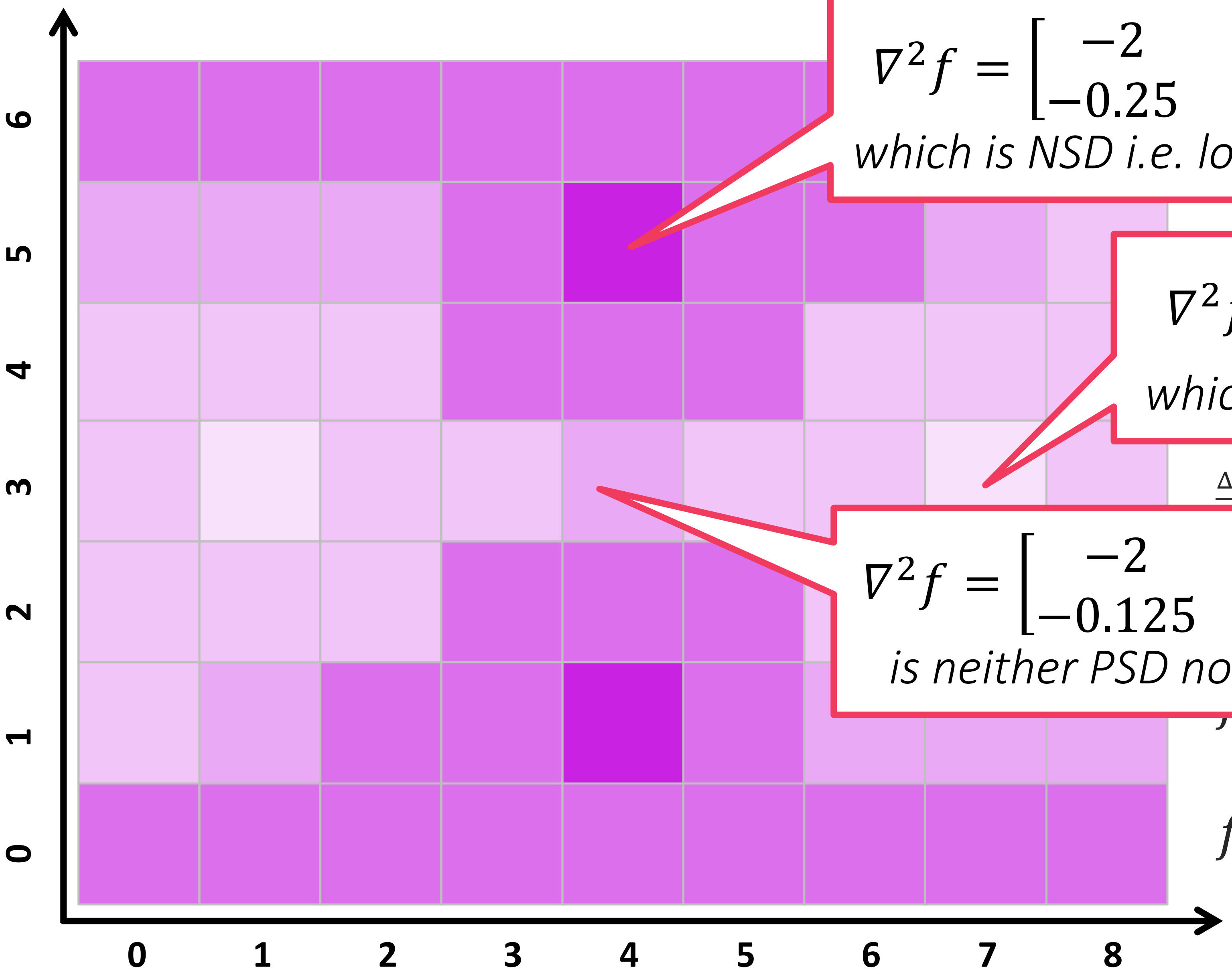
$$\frac{\Delta^2 f}{\Delta y^2} = f(x_0, y_0 + 1) + f(x_0, y_0 - 1) - 2f(x_0, y_0)$$

$$\frac{\Delta^2 f}{\Delta x \Delta y} = \frac{(f_{xy} + f_{yx})}{2} \text{ where}$$

$$f_{xy} = \frac{\frac{\Delta f}{\Delta x}(x_0, y_0 + 1) - \frac{\Delta f}{\Delta x}(x_0, y_0 - 1)}{2}$$

$$f_{yx} = \frac{\frac{\Delta f}{\Delta y}(x_0 + 1, y_0) - \frac{\Delta f}{\Delta y}(x_0 - 1, y_0)}{2}$$

A Toy Example – Hessians



$\nabla^2 f = \begin{bmatrix} -2 & -0.25 \\ -0.25 & -2 \end{bmatrix}$ *which is NSD i.e. local max*

$\nabla^2 f = \begin{bmatrix} 2 & 0.125 \\ 0.125 & 2 \end{bmatrix}$ *which is PSD i.e. local min*

$\nabla^2 f = \begin{bmatrix} -2 & -0.125 \\ -0.125 & 2 \end{bmatrix}$ *which is neither PSD nor NSD (saddle)*

...crete toy example, we can ... Hessian at (x_0, y_0) as

$$\begin{bmatrix} \frac{\Delta^2 f}{\Delta x^2} & \frac{\Delta^2 f}{\Delta x \Delta y} \\ \cdots & \frac{\Delta^2 f}{\Delta y^2} \end{bmatrix}$$ where

$$\ldots) - 2f(x_0, y_0)$$

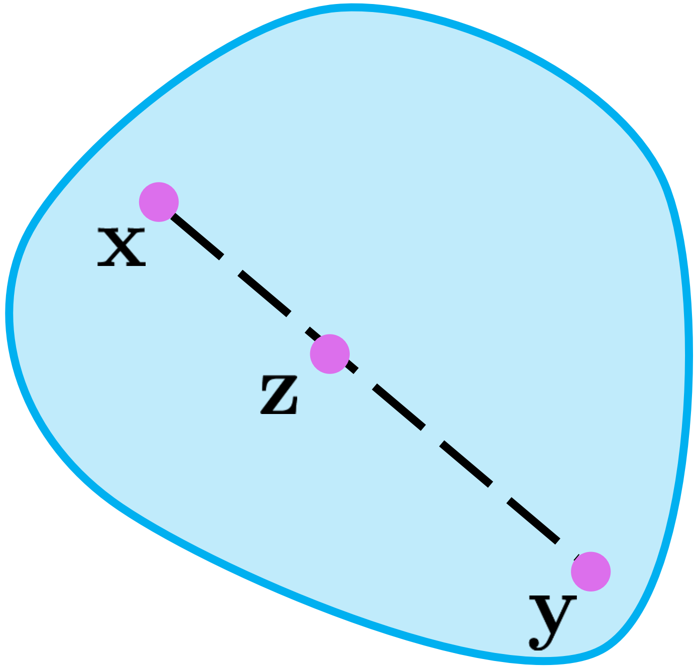
$$\frac{\Delta^2 f}{\Delta y^2} = f(x_0, y_0 + 1) + f(x_0, y_0 - 1) - 2f(x_0, y_0)$$

$$f_{xy} = \frac{\cdots - \frac{\Delta f}{\Delta x}(x_0, y_0 - 1)}{2}$$

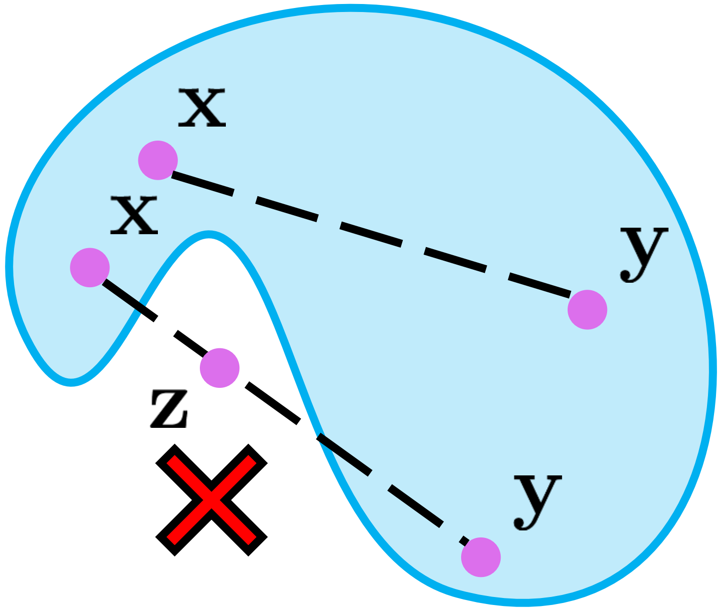
$$f_{yx} = \frac{\frac{\Delta f}{\Delta y}(x_0 + 1, y_0) - \frac{\Delta f}{\Delta y}(x_0 - 1, y_0)}{2}$$

Convex Sets

$$\mathcal{C} \subseteq \mathbb{R}^d$$



CONVEX SET

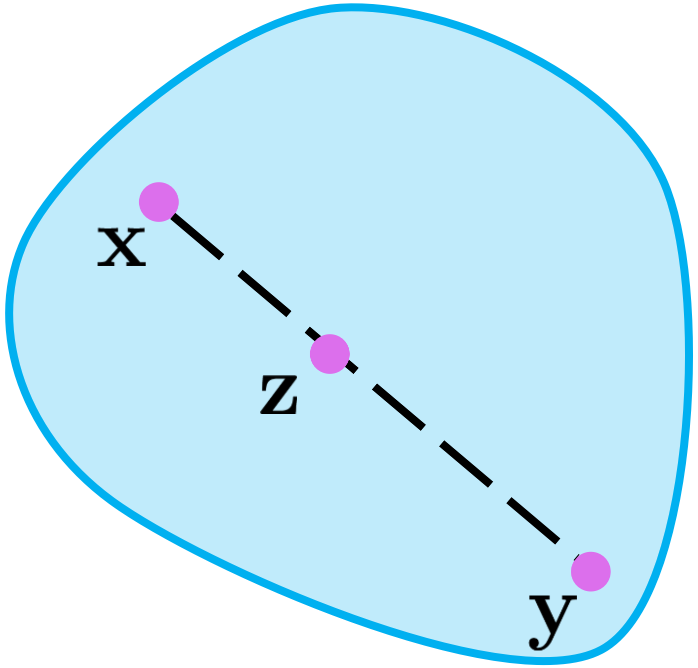


NON-CONVEX SET

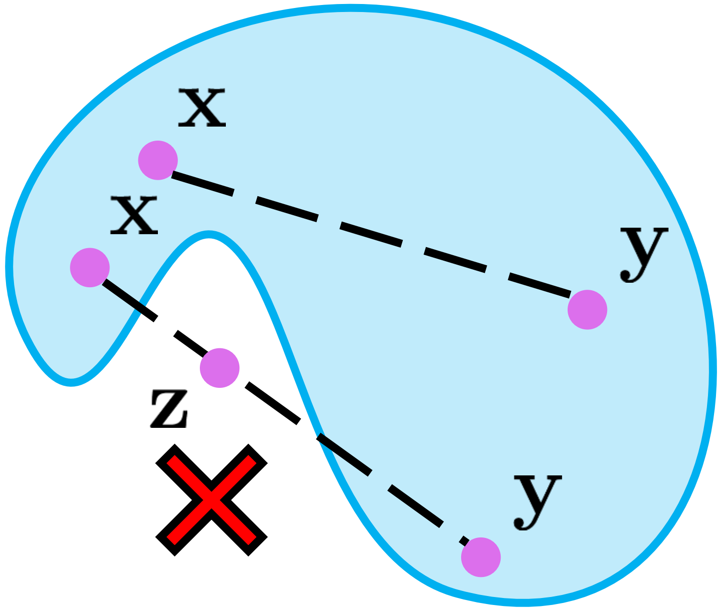
$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{C}$$
$$\forall \lambda \in [0, 1]$$
$$\mathbf{z} = \lambda \cdot \mathbf{x} + (1 - \lambda) \cdot \mathbf{y} \in \mathcal{C}$$

Convex Sets

$$\mathcal{C} \subseteq \mathbb{R}^d$$



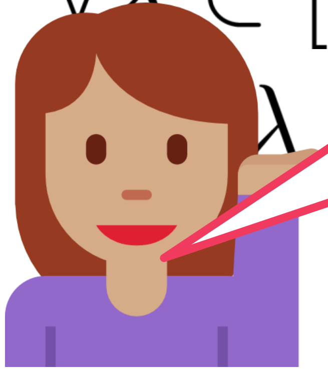
CONVEX SET



NON-CONVEX SET

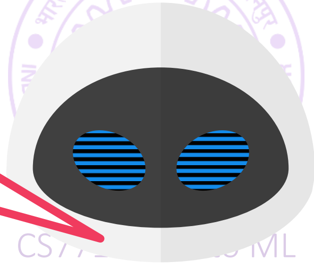
$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{C}$$

$$\forall \lambda \in [0,$$



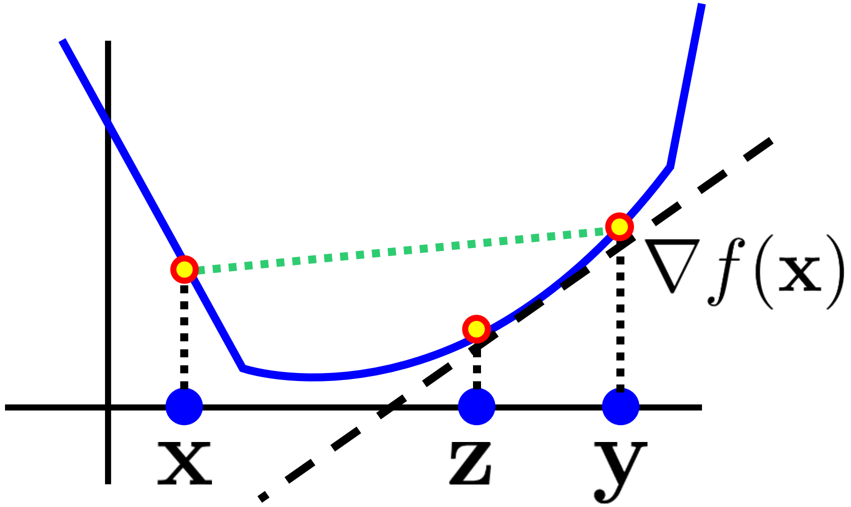
Think about which common shapes/objects are convex and which are not – balls, cuboids, stars, rectangles?

The intersection of two convex sets is always convex. The union may or may not be convex!

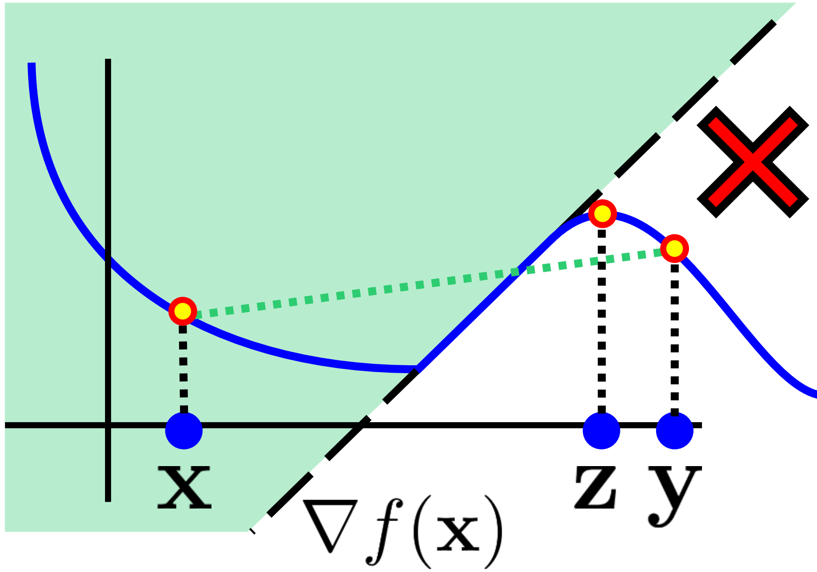


Convex Functions

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



CONVEX FUNCTION



NON-CONVEX FUNCTION

$\forall \mathbf{x}, \mathbf{y}$

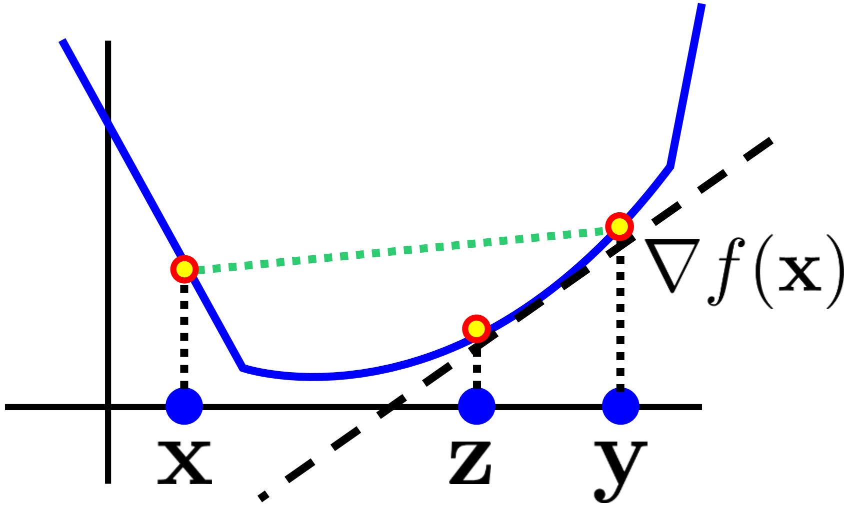
$\forall \lambda \in [0, 1]$

$$\mathbf{z} = \lambda \cdot \mathbf{x} + (1 - \lambda) \cdot \mathbf{y}$$

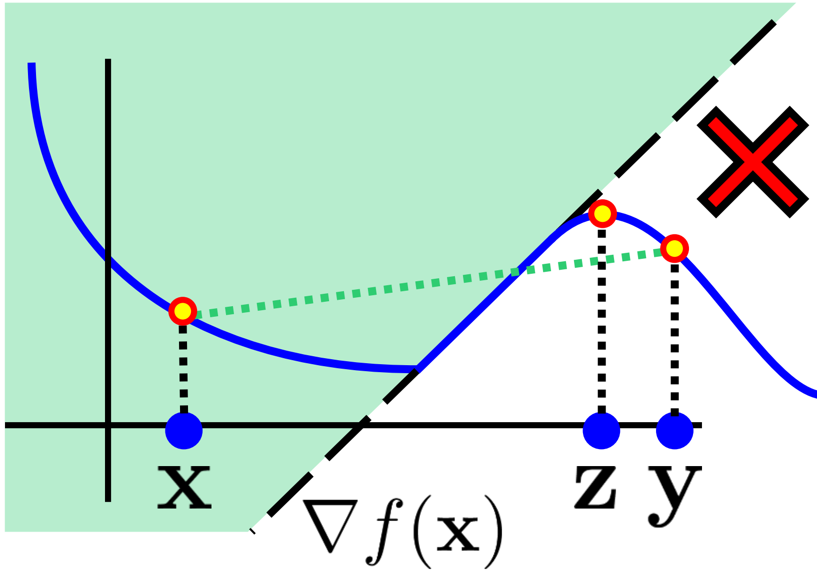
$$f(\mathbf{z}) \leq \lambda \cdot f(\mathbf{x}) + (1 - \lambda) \cdot f(\mathbf{y})$$

Convex Functions

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



CONVEX FUNCTION



NON-CONVEX FUNCTION

$\forall \mathbf{x}, \mathbf{y}$

$\forall \lambda \in [0, 1]$

$$\mathbf{z} = \lambda \cdot \mathbf{x} + (1 - \lambda) \cdot \mathbf{y}$$

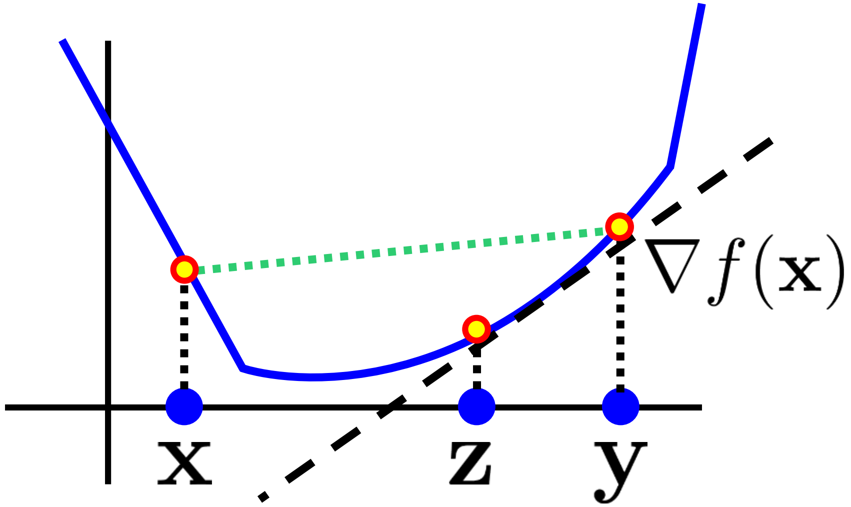
$$f(\mathbf{z}) \leq \lambda \cdot f(\mathbf{x}) + (1 - \lambda)$$

A convex function must lie below all its *chords*

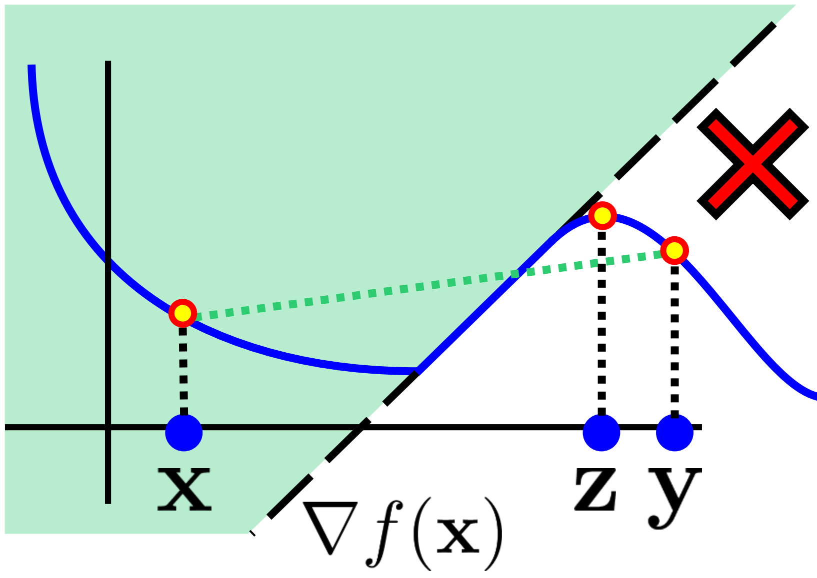
Convex Functions

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

$\forall \mathbf{x}, \mathbf{y}$



CONVEX FUNCTION



NON-CONVEX FUNCTION

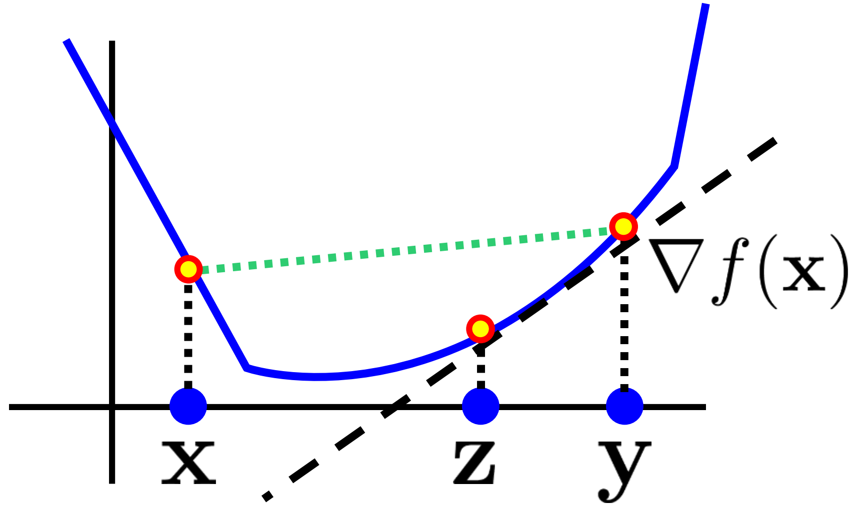
For differentiable functions, a nicer definition

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$$

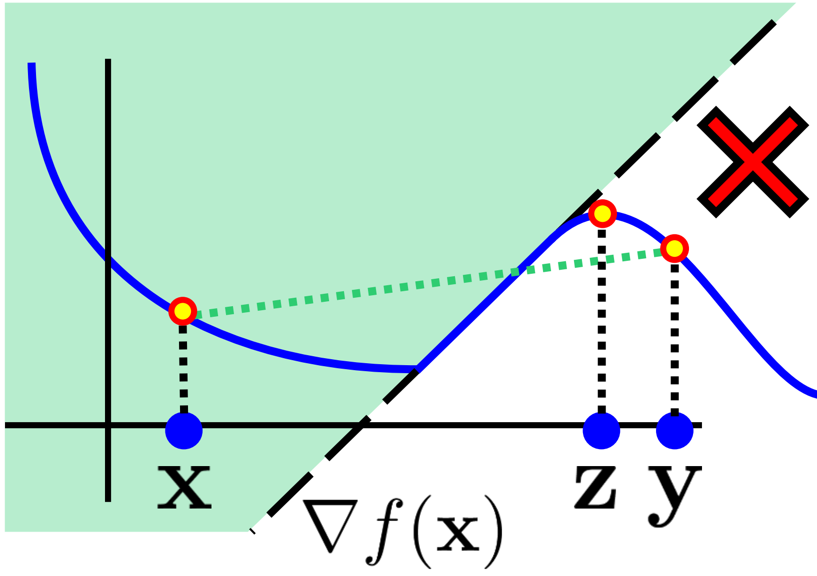
Convex Functions

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

$\forall \mathbf{x}, \mathbf{y}$



CONVEX FUNCTION



NON-CONVEX

TION

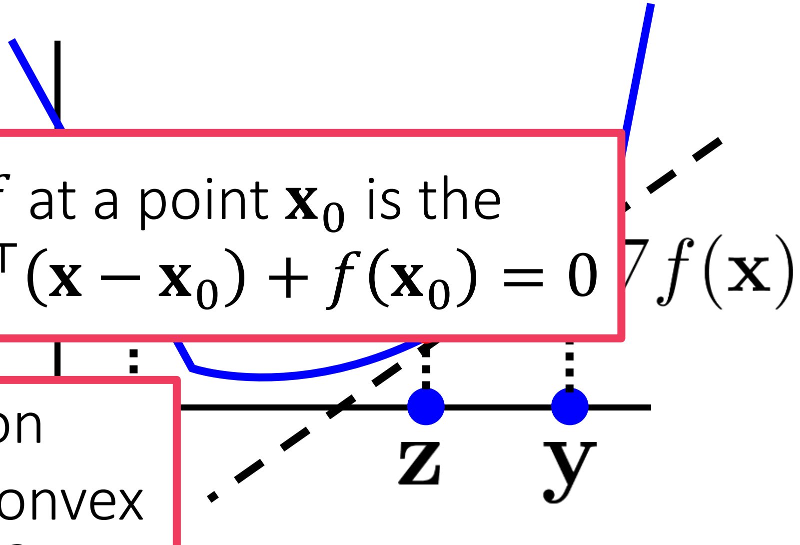
For differentiable functions, a ni

A differentiable convex function must lie above all its *tangents*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$$

Convex Functions

The tangent to f at a point $\mathbf{x}_0$ is the hyperplane $\nabla f(\mathbf{x}_0)^\top(\mathbf{x}-\mathbf{x}_0)+f(\mathbf{x}_0)=0$

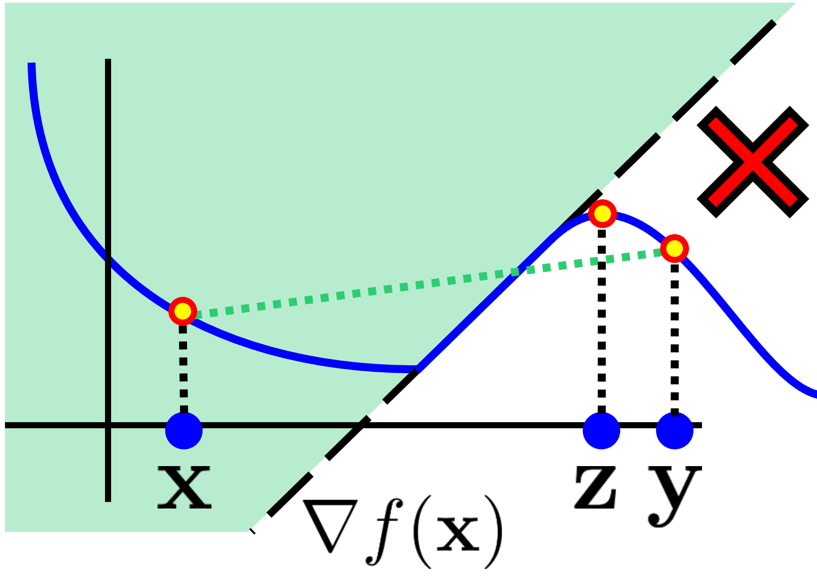


Think of common functions that are convex

1D examples: x^2

d-dim example: $\|\mathbf{x}\|_2^2$

EX FUNCTION



NON-CONVEX FUNCTION

, a nicer definition

The sum of two convex functions is always convex. The difference may or may not be convex

In fact a third definition exists for twice differentiable convex functions: their Hessian $\nabla^2 f(\mathbf{x})$ must be PSD everywhere

Checking for Convexity

All constant functions $f(\mathbf{x}) = c$ are convex

All linear functions $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$ are convex

Sums of convex functions are convex

Positive multiples of convex functions $c \cdot f(\mathbf{x}), c \geq \mathbf{0}$ are convex

If $g: \mathbb{R}^d \to \mathbb{R}$ is convex and $f: \mathbb{R} \to \mathbb{R}$ is convex and non-decreasing i.e. $a \geq b \Rightarrow f(a) \geq f(b)$, then $f \circ g: \mathbb{R}^d \to \mathbb{R}$ is convex

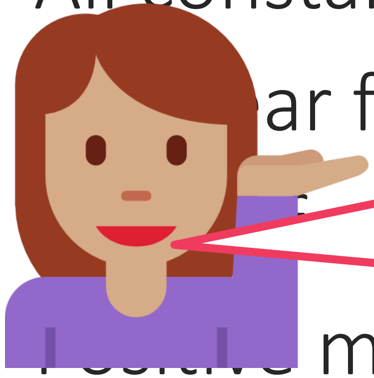
The Euclidean distance is convex $f(\mathbf{x}) = \|\mathbf{x}\|_2$ is convex

If $f: \mathbb{R} \to \mathbb{R}$ is convex then $g(\mathbf{x}) = f(\mathbf{a}^\top \mathbf{x} + b)$ is also convex

Checking for Convexity

All constant functions $f(\mathbf{x}) = c$ are convex

...ar fu... ...vex

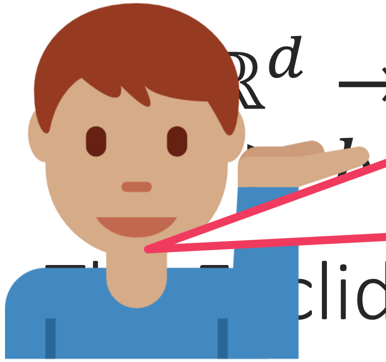


Many popular functions are concave e.g. $\log x$, $\sqrt{x}$. The negative of a concave function is always convex

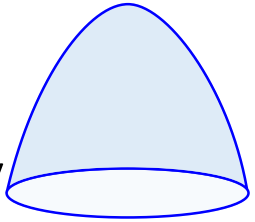
...multiples of convex functions $c \cdot f(\mathbf{x}), c \geq 0$ are convex

$\mathbb{R}^d \rightarrow$ ... convex and non-decreasing

... $\mathbb{R}^d \rightarrow \mathbb{R}$ is convex



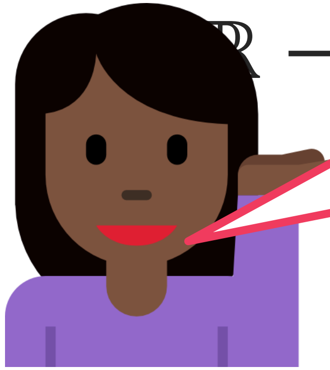
The negative of a convex function $-f(x)$ is called a *concave* function and they look like inverted cups



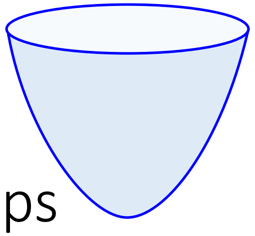
I also love concave functions since all local maxima are global maxima for a concave function

...clidean distance is convex $f(\mathbf{x}) =$ ...

$\mathbb{R} \rightarrow$ ... $g(\mathbf{x}) = f(\mathbf{a}^\top \mathbf{x} + b)$ is also co...



Convex functions look like cups



I love convex functions since all local minima are global minima for a convex function