Explanatory Notes for 6.390

Shaunticlair Ruiz (Current TA)

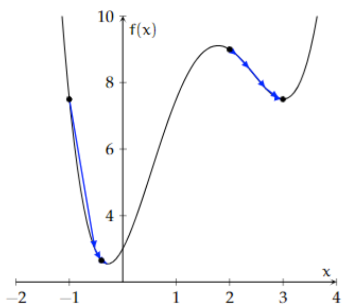
Fall 2022

Local minima

Even if we don't have that problem, we have a **different** one:

Gradient descent **gradually** improves our solution until it reaches one it's **satisfied** with. But, what if there are **multiple** solutions we could reach?

Are they all equally good?



Depending on your starting position (**initialization**), you could find a different local minimum!

Maybe not! So, if our function isn't **always convex**, we can end up with **multiple** "valleys", or **local** minima.

Definition 1

A **global** minimum is the **lowest** point on our entire function: the one with the lowest **output**.

A **local** minimum is one that is the **lowest** point among those points that are **near** it.

- For **local minima**, if you add or subtract a **small** amount ϵ, the value will **increase**.

So, we **won't** necessarily end up with the **global** minimum, even with a *small* η.

This shows that **initialization matters**!

Definition 2

Initialization is our "starting point": when we first **start** our algorithm, what are our **parameters** set to?

If we have a **different** starting position, we can find a **different** local minimum.

Concept 3

Gradient descent finds **local** minima near the initialization, not **global** minima.

This means, if our function has **multiple local minima** (not fully convex), our **initialization** can affect our **solution**.