Data Science

Introduction to Machine Learning: Gradient Descent

June 30, 2021

Recap on the general problem

Many Machine Learning problems take the following form:

$$\text{minimize}_\theta \sum_{i=1}^{m} l(h_\theta(x^{(i)}), y^{(i)})$$

One Approach

Gradient Descent!

One Approach

Gradient Descent!

1. The term gradient comes from calculus (a vector of partial derivatives)

One Approach

Gradient Descent!

1. The term gradient comes from calculus (a vector of partial derivatives)
2. We can ‘ride’ the gradient to some minimum (or maximum)

One Approach

Gradient Descent is search! The basic algorithm:

One Approach

Gradient Descent is search! The basic algorithm:

1. pick a starting point

One Approach

Gradient Descent is search! The basic algorithm:

1. pick a starting point
2. compute the sum of the loss over learning set

One Approach

Gradient Descent is search! The basic algorithm:

1. pick a starting point
2. compute the sum of the loss over learning set
3. compute the sum of the loss for points ‘nearby’

One Approach

Gradient Descent is search! The basic algorithm:

1. pick a starting point
2. compute the sum of the loss over learning set
3. compute the sum of the loss for points ‘nearby’
4. pick new parameters based on the gradient from the previous steps

One Approach

Gradient Descent is search! The basic algorithm:

1. pick a starting point
2. compute the sum of the loss over learning set
3. compute the sum of the loss for points 'nearby'
4. pick new parameters based on the gradient from the previous steps
5. repeat

One Approach

Gradient Descent is search! The basic algorithm:

1. pick a starting point
2. compute the sum of the loss over learning set
3. compute the sum of the loss for points ‘nearby’
4. pick new parameters based on the gradient from the previous steps
5. repeat
6. When do we stop?

One Approach

Gradient Descent is search! The basic algorithm:

1. pick a starting point
2. compute the sum of the loss over learning set
3. compute the sum of the loss for points 'nearby'
4. pick new parameters based on the gradient from the previous steps
5. repeat
6. When do we stop?
7. What assumptions have we baked in?

Gradient Descent

Assumptions

Gradient Descent

Assumptions

1. That the loss function has a gradient!

Gradient Descent

Assumptions

1. That the loss function has a gradient!
2. That there's only one minimum or maximum (not always true!)

Loss Functions

What we want:

Loss Functions

What we want:

1. Continuity

Loss Functions

What we want:

1. Continuity
2. Global minimum

Loss Functions

What we want:

1. Continuity
2. Global minimum
3. Cheap

Loss Functions

What we want:

1. Continuity
2. Global minimum
3. Cheap
4. Convex