Gradient Descent

Hamid Dehghani

School of Computer Science

Birmingham

January 2024

UNIVERSITY OF **BIRMINGHAM**

Intended Learning Outcome

- Understand why Gradient Descent is useful
- Understand and explain the concept of Gradient Descent
- Know what non-linear models are
- Be able to implement a simple regression algorithm

UNIVERSITY OF BIRMINGHAM

Gradient Descent

- An optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent
  - defined by the negative of the gradient.
- In machine learning, we use gradient descent to update the parameters of our model.
 - coefficients in Linear Regression and weights in neural networks.

Gradient Descent

- Consider the 3D graph in the context of a cost function.

- Aim is to move from the mountain in the top right corner (high cost) to the dark blue sea in the bottom left (low cost).

- The arrows represent the direction of steepest descent (negative gradient) from any given point–the direction that decreases the cost function as quickly as possible.
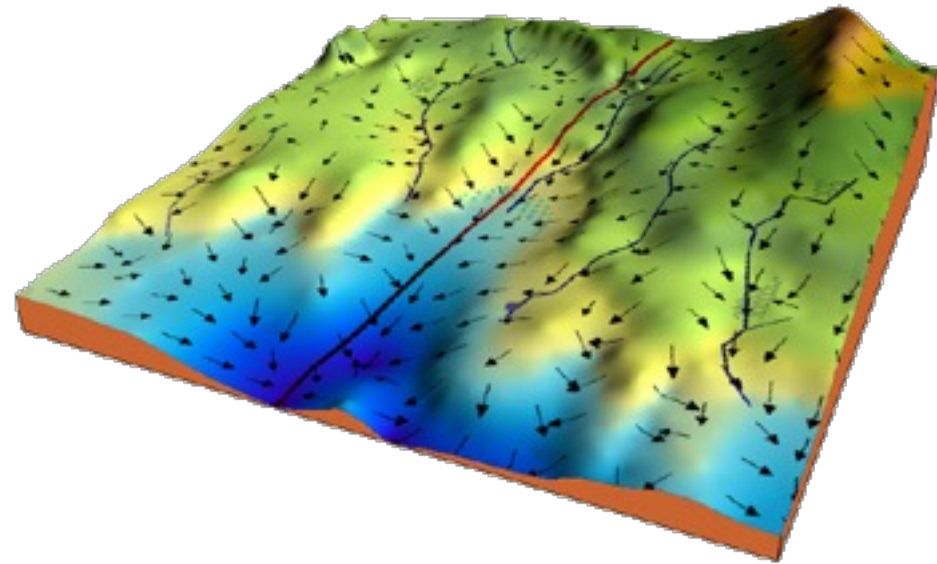
Basic Concept

- Starting at the top take first step downhill in the direction specified by the negative gradient.

- Recalculate the negative gradient (passing in the coordinates of our new point) and take another step in the direction it specifies.

- Continue this process iteratively until we get to the bottom of our graph, or to a point where we can no longer move downhill–a local minimum

UNIVERSITY OF BIRMINGHAM
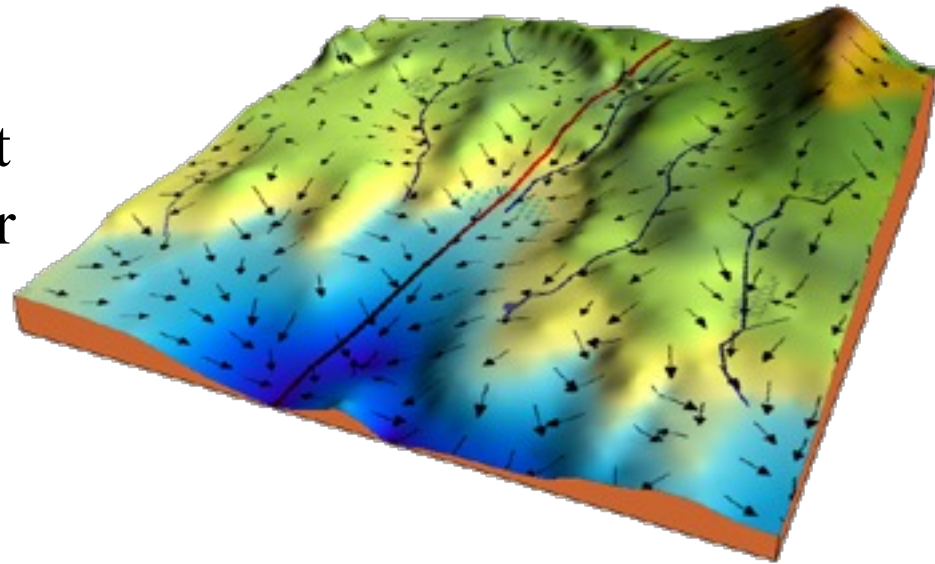
Learning rate

- Size of these steps is called the learning rate.
  - A high learning rate covers more ground each step, but risks overshooting the lowest point since the slope of the hill is constantly changing.
  - A very low learning rate, we can confidently move in the direction of the negative gradient since we are recalculating it so frequently.

- A low learning rate is more precise
  - but calculating the gradient is time-consuming, so it will take us a very long time to get to the bottom.
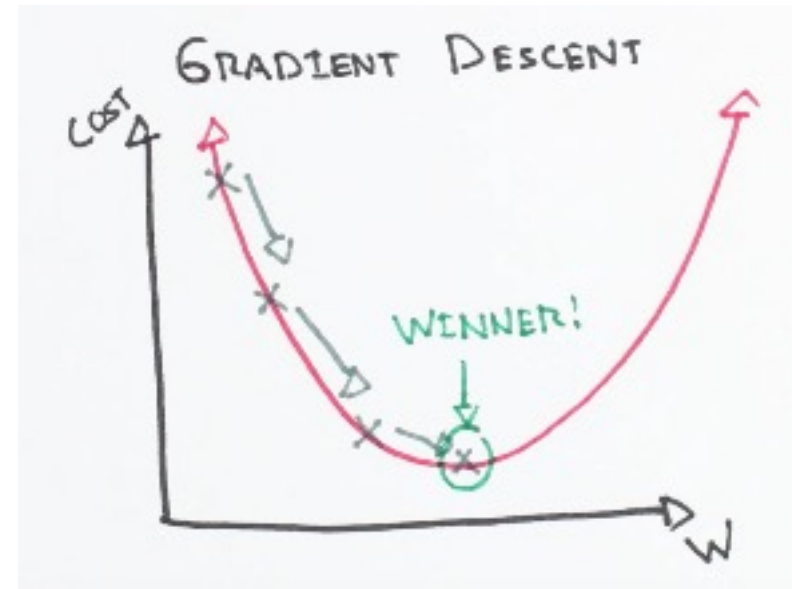
Cost function

- A Loss Functions tells us "how good" our model is at making predictions for a given set of parameters.

  - Refer to slides on Regression

    - Least Square Error

  - Refer to regularisation

    - Regularised Least Squares