Optimization

Tuan Nguyen

Ngày 16 tháng 1 năm 2021

Overview

What is Machine Learning?

Hình 1: Machine Learning

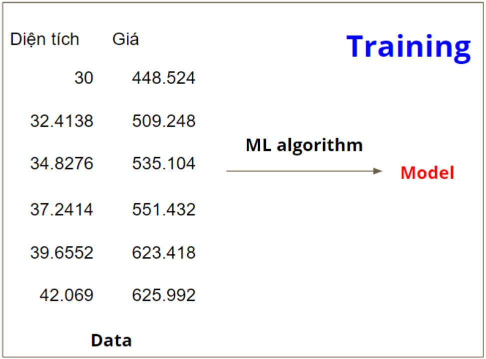
What is Machine Learning? (cont.)

Hình 2: Machine Learning

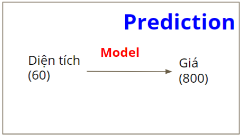
There are two main steps in Machine Learning task

- Training: Data -> Model
- Prediction: Model -> Predict

House price prediction

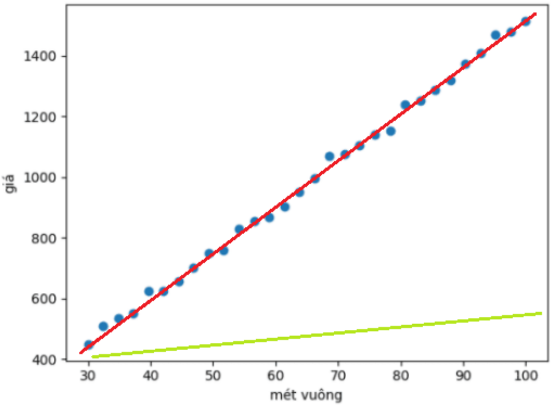


Hình 3: Training

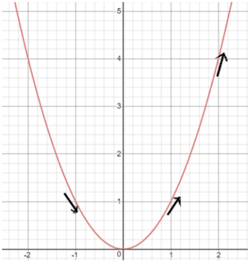


Hình 4: Prediction

Why optimization?



Hình 5: Choose model



Hình 6: Function f(x) = x^2

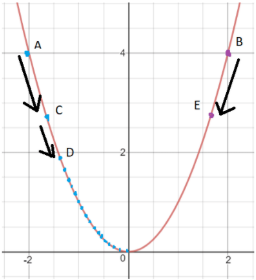
Function f(x) = x^2 $\rightarrow$ f'(x) = 2x. Remarks:

- f'(1) = 2 * 1 < f'(2) = 2 * 2 => the function at x = 2 is steeper than the function at x = 1 => the higher absolute value of gradient, the steeper function is.
- f'(-1) = 2 * (-1) = -2 < 0 => the function decreases (x increases, y decreases)

Gradient descent (cont.)

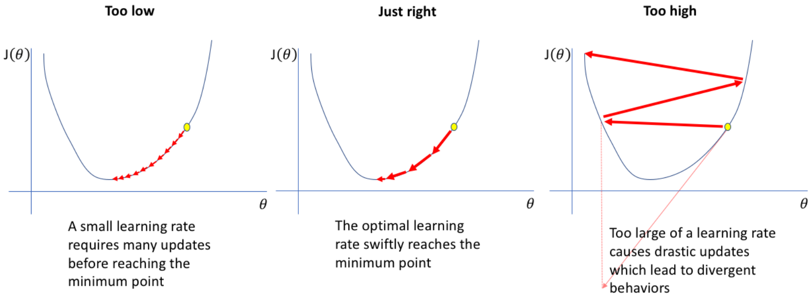
Steps to optimize the function f(x), $R \rightarrow R$, $x \rightarrow f(x)$:

1. Initialize randomly x = x_0
2. Update x = x - learning_rate × f'(x), learning_rate is a positive small number.
3. If f(x) is small enough, stop the algorithm. Otherwise, repeat the second step.



Hình 7: Gradient descent update

How to choose learning rate?



Hình 8: How to choose learning rate

Steps to optimize the function f(x), $R^n \rightarrow R$, $x \rightarrow f(x)$:

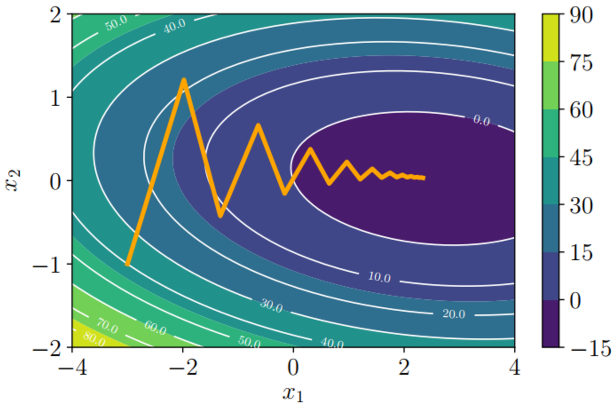
1. Initialize randomly x
2. Update x = x - learning_rate $\times$ $(\frac{df}{dx})^T$, learning_rate is a positive small number.
3. If f(x) is small enough, stop the algorithm. Otherwise, repeat the second step.

For example, f(x): $R^2 \rightarrow R$

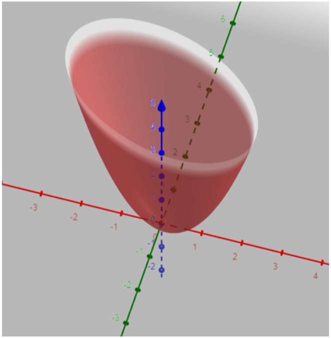
$$f(x) = \frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 20 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 5 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Initial $x_0 = \begin{bmatrix} -3 \\ -1 \end{bmatrix}$, iterate 5 steps of gradient descent algorithm.

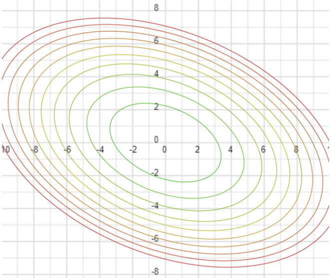
Gradient descent (cont.)



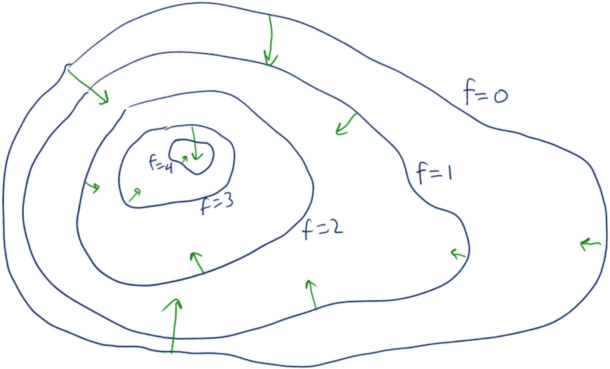
Hình 9: Gradient descent algorithm



Hình 10: 3D function
$f(x) = x^2 + 2y^2 + xy$



Hình 11: Countour plot



Hình 12: Directional gradient

Directional Derivatives

Lagrange multipliers

Lagrange multiplier technique lets you find the maximum or minimum of a multivariable function f(x, y,...) with the constrain:

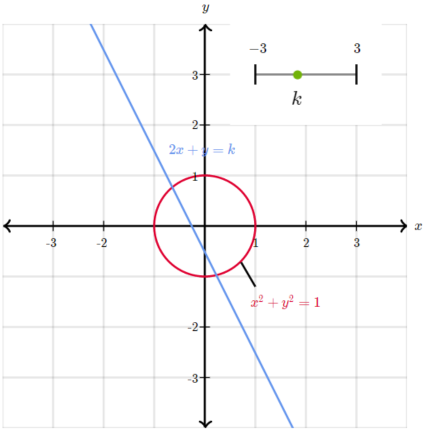
$$g(x, y, ...) = c$$

g is another multivariable function with the same input space as f, and c is a constant.

For example: Find the maximum and minimum of function $f(x, y) = 2x + y$, with a constrain:

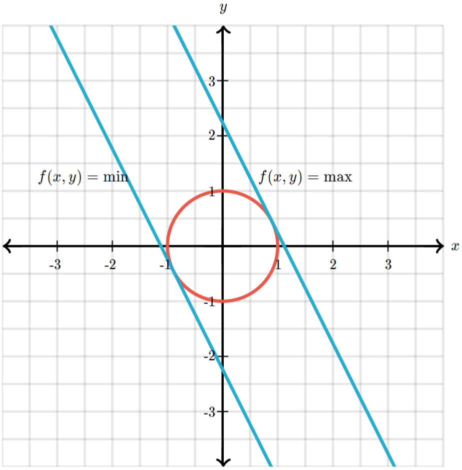
$$x^2 + y^2 = 1$$

Lagrange multipliers (cont.)

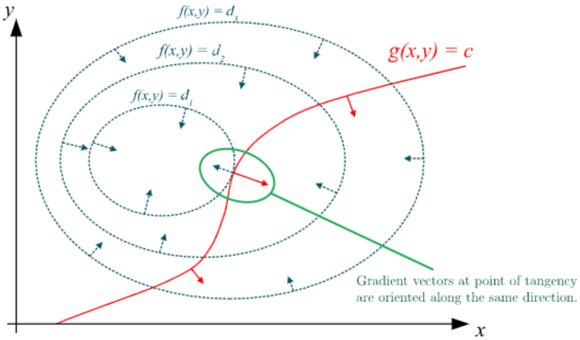


Hình 13: Constrained optimization

Lagrange multipliers (cont.)

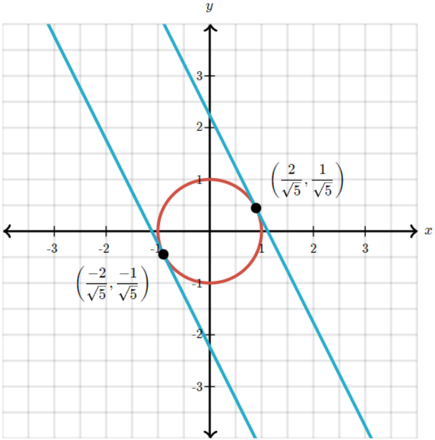


Hình 14: Maximum and minimum values of f



Hình 15: Tangent of f and g

At the tangent point (x_0, y_0), f'$(x_0, y_0) = \lambda$ g'(x_0, y_0) and we know that $x_0^2 + y_0^2 = 1 \Rightarrow$ find x_0, y_0



Hình 16: Maximum and minimum values of f

Equality Constrained Optimization

The steps to find optimal for f(x), $x \in R^n$, subject to:

$$h_i(x) = 0 \ \forall i = 1, 2, ..., m$$

- Introduce new variable $\lambda_1, \lambda_2, ..., \lambda_m$ and new function

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

- Set the gradient equal 0

$$\bigtriangledown \mathcal{L}(x, \lambda_1, \lambda_2, ..., \lambda_m) = 0$$

- For each solution in step 2, find the minimum and maximum