

Machine Learning for Engineering and Science Applications

Optimization-1

Unconstrained Optimization

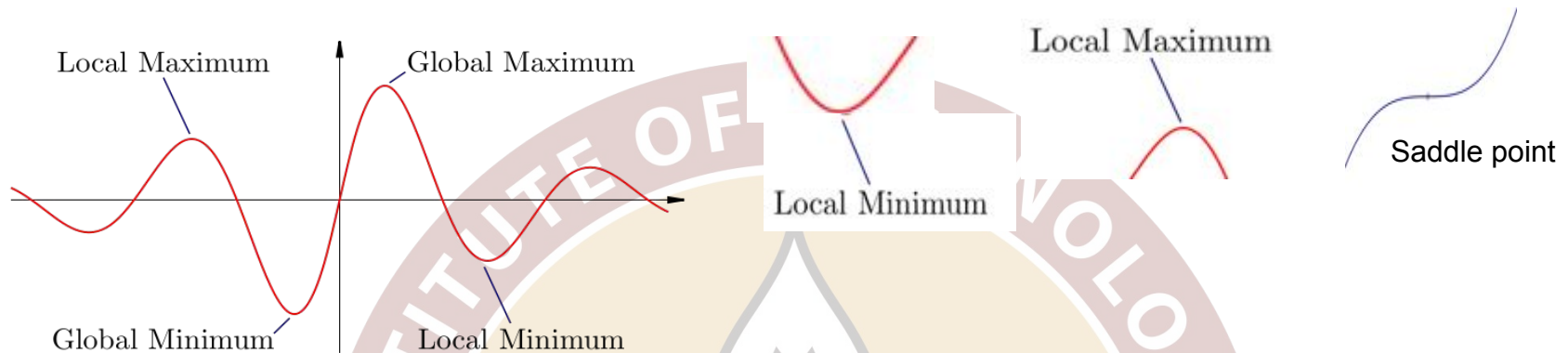
Optimization and Machine Learning

- The basic idea behind much of Machine learning is to build models that map input data to output data.
- We derive parameters of these models based on “training” on some given data.
- We can see the whole of machine learning as a process that finds the **best/optimal Model for the given data.**
- Most machine learning problems are, therefore, ultimately optimization problems
- We will review and introduce some optimization techniques in the coming videos.

Optimization

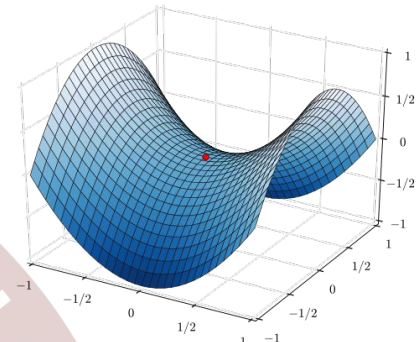
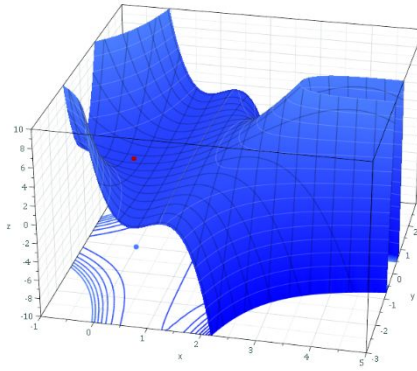
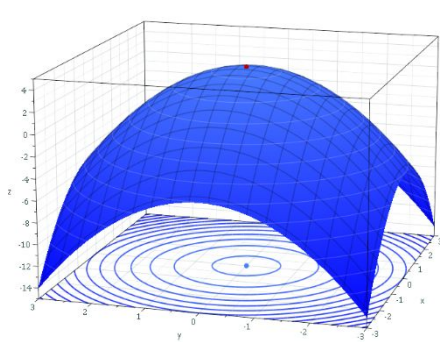
- The general optimization task is to maximize or minimize a function $f(\mathbf{x})$ by varying \mathbf{x} .
 - The function $f(\mathbf{x})$ is called the **objective function** or **cost function** or **loss function**
 - The function $f(\mathbf{x})$ maybe a scalar (single objective) or a vector (multi-objective)
 - In this course (and most of Machine Learning) we deal only with a single objective. That is, $f(\mathbf{x})$ is a scalar.
 - However, \mathbf{x} is, in general, a vector.
 - Therefore, $f: \mathbb{R}^n \rightarrow \mathbb{R}$
 - For example, $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$. Here, $f: \mathbb{R}^3 \rightarrow \mathbb{R}$
- It is possible to reduce all optimization problems to minimization problems.
 - That is, all problems can be written as find \mathbf{x} that minimizes $f(\mathbf{x})$
 - Any maximization problem can be written as minimization of $-f(\mathbf{x})$
- We denote the solution to the problem as $\mathbf{x}^* = \arg \min f(\mathbf{x})$

Optimization – Scalar x



- We will look at the **unconstrained problem**. That is, find x that minimizes $f(x)$ with $x \in \mathbb{R}$. That is, no constraints on x .
- It can be shown that any local extremum will have the property $f'(x) = 0$
 - Such points are called **stationary points** or **critical points**.
 - The stationary point may be a (local) minimum, maximum or saddle point
- If $f''(x) > 0$, it is a local minimum
- If $f''(x) < 0$, it is a local maximum
- If $f''(x) = 0$, it could be a saddle point
- The absolute lowest/highest level of $f(x)$ is called the global maximum/minimum

Optimization – Multivariate x



- In this case the **unconstrained optimization problem** is to find \mathbf{x} that minimizes $f(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^n$. That is, there are no constraints on \mathbf{x} .
- Since \mathbf{x} is now a vector quantity, we need to evaluate the **gradient** $\frac{\partial f}{\partial \mathbf{x}} \equiv \nabla_{\mathbf{x}} f$
 - It can be shown that any local extremum will have the property $\nabla_{\mathbf{x}} f = \mathbf{0}$
 - Such points are called **stationary points** or **critical points**.
 - The stationary point may be a (local) minimum, maximum or saddle point
- The type of critical point is decided by the nature of the **Hessian** $H_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$
- If $H_{i,j}$ is positive definite it is a local minimum
- If $H_{i,j}$ is negative definite it is a local maximum
- If $H_{i,j}$ is indefinite (i.e. neither p.d or n.d) then it is a saddle point