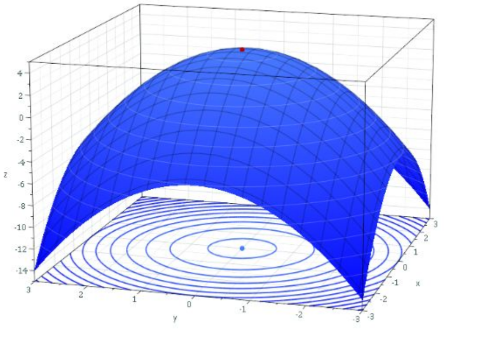
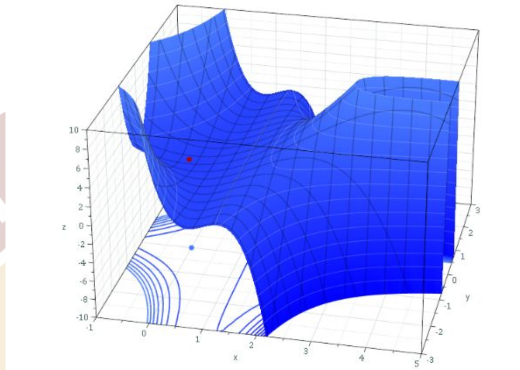
Machine Learning for Engineering and Science Applications

Numerical Optimization
Gradient Descent

Need for Numerical Optimization





- Optimization we saw so far was analytical.
- This requires explicit expressions for the objective function in terms of the features (variables).
  - Example : $J(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 + 4$
- However, usually we only know the function as a “black” box.
  - In machine learning this “black box” is our Machine Learning Model (e.g. Neural network)
- So, we have to develop numerical (rather than analytical techniques)

Iterative optimization -- Fundamental idea

- We want to drive $\nabla_x J$ to 0 but we do not have an analytical expression.
- Guess for x
- However, usually we only know the function as a “black” box.
  - In machine learning this “black box” is our Machine Learning Model (e.g. Neural network)
- So, we have to develop numerical (rather than analytical techniques)

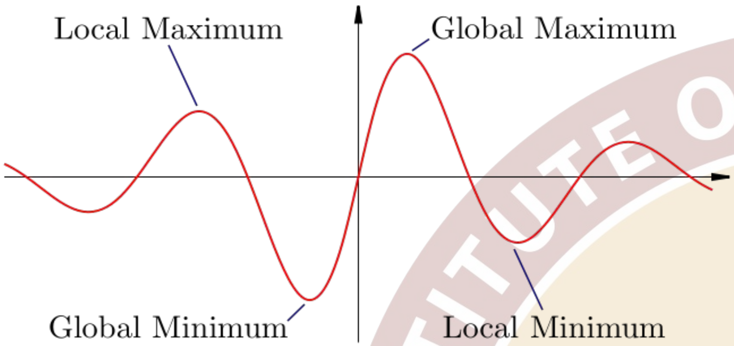
Optimization

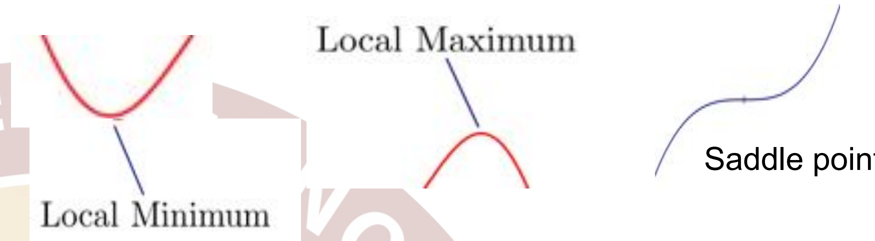
- The general optimization task is to maximize or minimize a function $f(\mathrm{x})$ by varying x.
  - The function $f(\mathrm{x})$ is called the objective function or cost function or loss function
  - The function $f(\mathrm{x})$ maybe a scalar (single objective) or a vector (multi-objective)
  - In this course (and most of Machine Learning) we deal only with a single objective. That is, $f(\mathrm{x})$ is a scalar.
  - However, 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 x that minimizes $f(\mathrm{x})$
  - Any maximization problem can be written as minimization of $-f(\mathrm{x})$

- We denote the solution to the problem as $\mathbf{x}^* = \arg\min f(\mathbf{x})$

Optimization – Scalar x



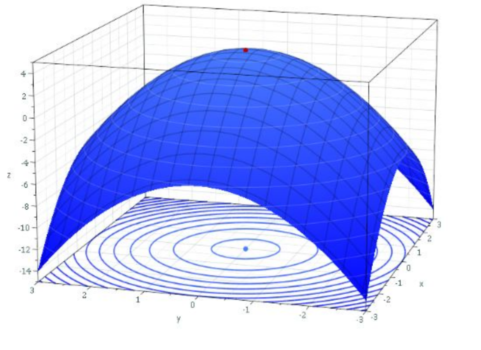


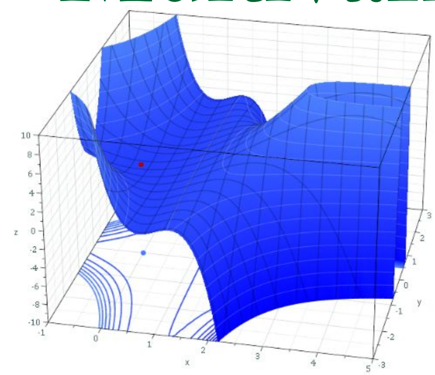
- We will look at the **unconstrained problem**. That is, find x that minimizes $f(x)$ with $x \in \mathbb{R}$. That is, no constraints on x.
- It can be shown that any local extremum will have the property $f'(x) = 0$
  - Such points are called **stationary points** or **critical points**.
  - The stationary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

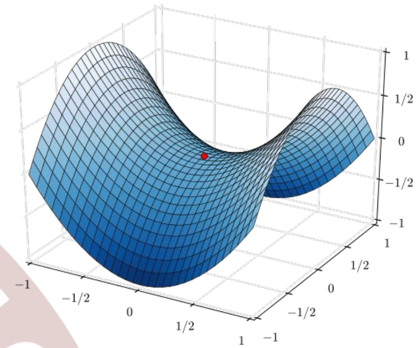
- If $f''(x) > 0$, it is a local minimum
- If $f''(x) < 0$, it is a local maximum
- If $f''(x) = 0$, it could be a saddle point
- The absolute lowest/highest level of $f(x)$ is called the global maximum/minimum

https://commons.wikimedia.org/wiki/File:Maxima_and_Minima.svg

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 x.
- Since $\mathbf{x}$ is now a vector quantity, we need to evaluate the gradient $\frac{\partial f}{\partial \mathbf{x}} \equiv \nabla_x f$
  - 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

https://en.wikipedia.org/wiki/Maxima_and_minima#/media/File:MaximumParaboloid.png