مبانی یادگیری ماشین

Intro to Machine Learning

بابک نجار اعرابی

دانشکده مهندسی برق و کامپیوتر دانشگاه تهران

نیم سال اول سال تحصیلی 1403-04

موضوع این جلسه

مروری بر روش های بهینه سازی

جلسه دوم

Babak Nadjar Araabi

School of Electrical & Computer Eng

University of Tehran

ECE-UT - Fall 2024

A very short review of Basic Math you need

Gradient

In vector calculus, the **gradient** of a scalar-valued differentiable function f of several variables is the vector field (or vector-valued function) ∇f whose value at a point p gives the direction and the rate of fastest increase.

When a coordinate system is used in which the basis vectors are not functions of position, the gradient is given by the vector[a] whose components are the partial derivatives of f at p.[2] That is, for $f\colon \mathbb{R}^n \to \mathbb{R}$, its gradient $\nabla f\colon \mathbb{R}^n \to \mathbb{R}^n$ is defined at the point $p = (x_1, \ldots, x_n)$ in n-dimensional space as the vector[b]

$$\nabla f(p) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(p) \\ \vdots \\ \frac{\partial f}{\partial x_n}(p) \end{bmatrix}.$$

Hessian

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a function taking as input a vector $\mathbf{x} \in \mathbb{R}^n$ and outputting a scalar $f(\mathbf{x}) \in \mathbb{R}$. If all second-order partial derivatives of f exist, then the Hessian matrix $\mathbf{H}$ of f is a square $n \times n$ matrix, usually defined and arranged as

$$\mathbf{H}_f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\[2.2ex] \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\[2.2ex] \vdots & \vdots & \ddots & \vdots \\[2.2ex] \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

That is, the entry of the ith row and the jth column is

$$(\mathbf{H}_f)_{i,j} = \frac{\partial^2 f}{\partial x_i\,\partial x_j}.$$

Hessian cont.

- If furthermore the second partial derivatives are all continuous, the Hessian matrix is a symmetric matrix by the symmetry of second derivatives.

Jacobian

Suppose $\mathbf{f}: \mathbf{R}^n \to \mathbf{R}^m$ is a function such that each of its first-order partial derivatives exists on $\mathbf{R}^n$. This function takes a point $\mathbf{x} \in \mathbf{R}^n$ as input and produces the vector $\mathbf{f}(\mathbf{x}) \in \mathbf{R}^m$ as output. Then the Jacobian matrix of $\mathbf{f}$, denoted $\mathbf{J_f} \in \mathbf{R}^{m \times n}$, is defined such that its $(i,j)^{\text{th}}$ entry is $\frac{\partial f_i}{\partial x_j}$, or explicitly

$$\mathbf{J_f} = \begin{bmatrix} \dfrac{\partial \mathbf{f}}{\partial x_1} & \cdots & \dfrac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^{\mathrm{T}} f_1 \\ \vdots \\ \nabla^{\mathrm{T}} f_m \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Jacobian cont.

The Hessian matrix of a function f is the transpose of the Jacobian matrix of the gradient of the function f; that is: $\mathbf{H}(f(\mathbf{x})) = \mathbf{J}(\nabla f(\mathbf{x}))^T$.

Positive (semi)Definite Matrix

An $n \times n$ symmetric real matrix M is said to be **positive-definite** if $\mathbf{x}^\top M \mathbf{x} > 0$ for all non-zero $\mathbf{x}$ in $\mathbb{R}^n$. Formally,

$$\boxed{M \text{ positive-definite} \quad \iff \quad \mathbf{x}^\top M\, \mathbf{x} > 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}}$$

An $n \times n$ symmetric real matrix M is said to be **positive-semidefinite** or **non-negative-definite** if $\mathbf{x}^\top M \mathbf{x} \geq 0$ for all $\mathbf{x}$ in $\mathbb{R}^n$. Formally,

$$\boxed{M \text{ positive semi-definite} \quad \iff \quad \mathbf{x}^\top M\, \mathbf{x} \geq 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n}$$

Negative (semi)Definite Matrix

An $n \times n$ symmetric real matrix M is said to be **negative-definite** if $\mathbf{x}^\top M \mathbf{x} < 0$ for all non-zero $\mathbf{x}$ in $\mathbb{R}^n$. Formally,

$$\boxed{M \text{ negative-definite} \quad \Longleftrightarrow \quad \mathbf{x}^\top M \mathbf{x} < 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}}$$

An $n \times n$ symmetric real matrix M is said to be **negative-semidefinite** or **non-positive-definite** if $\mathbf{x}^\top M \mathbf{x} \leq 0$ for all $\mathbf{x}$ in $\mathbb{R}^n$. Formally,

$$\boxed{M \text{ negative semi-definite} \quad \Longleftrightarrow \quad \mathbf{x}^\top M \mathbf{x} \leq 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n}$$

Linear (or Vector) Space

A vector space over a field F is a non-empty set V together with a binary operation and a binary function that satisfy the eight axioms listed below. In this context, the elements of V are commonly called *vectors*, and the elements of F are called *scalars*.[2]

- The binary operation, called *vector addition* or simply *addition* assigns to any two vectors $\mathbf{v}$ and $\mathbf{w}$ in V a third vector in V which is commonly written as $\mathbf{v} + \mathbf{w}$, and called the *sum* of these two vectors.
- The binary function, called *scalar multiplication*, assigns to any scalar a in F and any vector $\mathbf{v}$ in V another vector in V, which is denoted $a\mathbf{v}$.[nb 2]

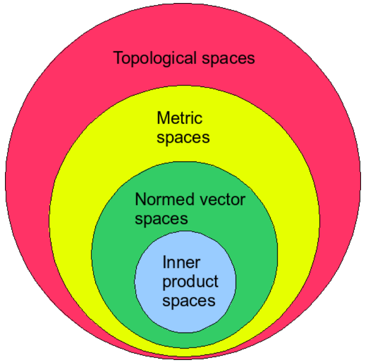
To have a vector space, the eight following axioms must be satisfied for every $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in V, and a and b in F.[3]

Linear (or Vector) Space cont.

Axiom	Statement
Associativity of vector addition	$\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
Commutativity of vector addition	$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
Identity element of vector addition	There exists an element $\mathbf{0} \in V$, called the *zero vector*, such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$.
Inverse elements of vector addition	For every $\mathbf{v} \in V$, there exists an element $-\mathbf{v} \in V$, called the *additive inverse* of $\mathbf{v}$, such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
Compatibility of scalar multiplication with field multiplication	$a(b\mathbf{v}) = (ab)\mathbf{v}$ [nb 3]
Identity element of scalar multiplication	$1\mathbf{v} = \mathbf{v}$, where 1 denotes the multiplicative identity in F.
Distributivity of scalar multiplication with respect to vector addition	$a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
Distributivity of scalar multiplication with respect to field addition	$(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$

Basis and Dimension of a Linear Space ++

- Linear combination
- Linear independence (of a set of vectors)
- Linear subspace
- Linear span (of a subset of vector space)
- Basis: A subset of a vector space is a basis if its elements are linearly independent and span the vector space
- Dimension
- Orthogonal Basis, Orthonormal Basis
- Generator set of vector space (or spanning set)
- Inner product, Norm, and Metric/Distance



Null Space & Nullity

Direct Sum

Tylor Expansion

The Taylor series of a real or complex-valued function $f(x)$, that is infinitely differentiable at a real or complex number a, is the power series

$$f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\frac{f'''(a)}{3!}(x-a)^3+\cdots=\sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(x-a)^n$$

- Maclaurin Expansion

$$f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+\frac{f'''(0)}{3!}x^3+\cdots=\sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{n!}x^n$$

Multivariate Taylor Expansion

$$T(x_1,\ldots,x_d) = \sum_{n_1=0}^{\infty}\cdots\sum_{n_d=0}^{\infty} \frac{(x_1-a_1)^{n_1}\cdots(x_d-a_d)^{n_d}}{n_1!\cdots n_d!}\left(\frac{\partial^{n_1+\cdots+n_d} f}{\partial x_1^{n_1}\cdots\partial x_d^{n_d}}\right)(a_1,\ldots,a_d)$$

$$= f(a_1,\ldots,a_d) + \sum_{j=1}^{d}\frac{\partial f(a_1,\ldots,a_d)}{\partial x_j}(x_j-a_j) + \frac{1}{2!}\sum_{j=1}^{d}\sum_{k=1}^{d}\frac{\partial^2 f(a_1,\ldots,a_d)}{\partial x_j \partial x_k}(x_j-a_j)(x_k-a_k)$$

$$+ \frac{1}{3!}\sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{l=1}^{d}\frac{\partial^3 f(a_1,\ldots,a_d)}{\partial x_j \partial x_k \partial x_l}(x_j-a_j)(x_k-a_k)(x_l-a_l) + \cdots$$

For example, for a function $f(x, y)$ that depends on two variables, x and y, the Taylor series to second order about the point (a, b) is

$$f(a,b) + (x-a)f_x(a,b) + (y-b)f_y(a,b) + \frac{1}{2!}\Big((x-a)^2 f_{xx}(a,b) + 2(x-a)(y-b)f_{xy}(a,b) + (y-b)^2 f_{yy}(a,b)\Big)$$

where the subscripts denote the respective partial derivatives.

Multivariate Taylor Expansion cont.

A second-order Taylor series expansion of a scalar-valued function of more than one variable can be written compactly as

$$T(\mathbf{x}) = f(\mathbf{a}) + (\mathbf{x}-\mathbf{a})^\mathsf{T} D f(\mathbf{a}) + \frac{1}{2!}(\mathbf{x}-\mathbf{a})^\mathsf{T} \left\{D^2 f(\mathbf{a})\right\} (\mathbf{x}-\mathbf{a}) + \cdots,$$

where $D f(\mathbf{a})$ is the gradient of f evaluated at $\mathbf{x} = \mathbf{a}$ and $D^2 f(\mathbf{a})$ is the Hessian matrix. Applying the multi-index notation the Taylor series for several variables becomes

$$T(\mathbf{x}) = \sum_{|\alpha|\geq 0} \frac{(\mathbf{x}-\mathbf{a})^\alpha}{\alpha!} \left(\partial^\alpha f\right)(\mathbf{a}),$$

which is to be understood as a still more abbreviated multi-index version of the first equation of this paragraph, with a full analogy to the single variable case.

Eigen Value and Eigen Vector

Convex Set

Convex Function

Feasible Solusion

Convex Optimization: Definition