Integration

AI503
Shan Shan

Lecture 12

Change of variables

Jacobian matrix

For $T(u, v) = (x(u, v), y(u, v))$ define

$$J(u, v) = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix}.$$

Provided the partial derivatives exist.

The Jacobian matrix of the linear map $T(u, v) = A \begin{pmatrix} u \\ v \end{pmatrix}$ is A.

Change of variables formula (2D)

Theorem

Let $T : D^* \to D$ be one-to-one and onto (up to boundaries). For integrable f on D:

$$\iint_D f(x, y)\, dx\, dy = \iint_{D^*} f\big(x(u, v), y(u, v)\big)\, |\det(J(u, v))|\, du\, dv.$$

Why the Jacobian?

Counterexample to naive formula without Jacobian:

Consider a linear map T that scales area by factor $c \neq 1$. Without the Jacobian area (and integrals) would be incorrect.

Polar coordinates

Apply the formula to $T(r, \theta) = (r\cos\theta, r\sin\theta)$. Compute

$$\det(J(r,\theta)) = \det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r.$$

Hence

$$\iint_D f(x,y)\,dx\,dy = \int_0^{2\pi}\int_0^1 f(r\cos\theta, r\sin\theta)\,r\,dr\,d\theta.$$

Gaussian integral (application)

Use polar coordinates to compute

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx.$$

Standard trick: square the integral and use polar coordinates to obtain $\sqrt{\pi}$.

Jacobian in 3D

For $T(u, v, w) = (x(u, v, w), y(u, v, w), z(u, v, w))$,

$$J = \frac{\partial(x, y, z)}{\partial(u, v, w)} = \begin{vmatrix} x_u & x_v & x_w \\ y_u & y_v & y_w \\ z_u & z_v & z_w \end{vmatrix}.$$

Change of variables formula (3D)

Let W and W^* be elementary region in $\mathbb{R}^3$ and let $T : W^* \to W$ be one-to-one and onto W (except possibly on a set that is the union of graphs of functions of two variables). Then for any integrable function $f : W \to \mathbb{R}$, we have

$$\begin{aligned}
&\iiint_W f(x, y, z)\, dV \\
&= \iiint_{W^*} f\left(x(u, v, w), y(u, v, w), z(u, v, w)\right)\, |J|\, du\, dv\, dw.
\end{aligned}$$

Cylindrical coordinates

Let $x = r\cos\theta$, $y = r\sin\theta$ and $z = z$. Let W^* consists of points in the form (r, θ, z) where $0 \leq r \leq 1, 0 \leq \theta \leq 2\pi, 0 \leq z \leq 1$. Find the image set and compute its Jacobian.

Spherical coordinates

Let $x = \rho \sin\phi \cos\theta$, $y = \rho \sin\phi \cos\theta$ and $z = \rho \cos\phi$. Compute its Jacobian and use this to derive the change of variable spherical cooridnate formula.

Final notes

- Check one-to-one condition or restrict domain when using change of variables.
- Always include absolute value of Jacobian.

Taylor polynomials

Motivation: Linear Approximation

- For a differentiable function $f(x)$, near $x = a$:

$$f(x) \approx f(a) + f'(a)(x - a)$$

- This is the **linear approximation** or **tangent line**.

Motivation: Linear Approximation

- For a differentiable function $f(x)$, near $x = a$:

$$f(x) \approx f(a) + f'(a)(x - a)$$

- This is the **linear approximation** or **tangent line**.

Example: $f(x) = \cos(x)$ at $a = 0$

$$\cos(x) \approx 1 - 0 \cdot x = 1$$

$$\text{Better: } \cos(x) \approx 1 - \frac{x^2}{2!}$$

Taylor Polynomials

Definition

The n-th degree Taylor polynomial of f at $x = a$ is

$$T_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Taylor Polynomials

Definition

The n-th degree Taylor polynomial of f at $x = a$ is

$$T_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

The idea: include higher-order derivatives to improve the local approximation.

Example: Taylor Polynomials for $\cos(x)$ around 0

$$T_0(x) = 1$$
$$T_2(x) = 1 - \frac{x^2}{2!}$$
$$T_4(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!}$$
$$T_6(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!}$$

$$\cos(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}$$

Visualization: Better Approximations

- $T_2(x)$ approximates well near 0
- $T_4(x)$ and $T_6(x)$ improve accuracy over larger intervals
- As $n \to \infty$, $T_n(x) \to \cos(x)$

Lagrange Error Bound for Taylor Polynomials

Suppose f has $(n+1)$ continuous derivatives on an interval containing a and x.

Taylor's Theorem with Remainder

$$f(x) = T_n(x) + R_n(x)$$

where

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$$

for some c between a and x. If $|f^{(n+1)}(t)| \leq M$ on the interval, then

$$|R_n(x)| \leq \frac{M}{(n+1)!}|x-a|^{n+1}.$$

- This quantifies how fast $T_n(x)$ converges to $f(x)$.
- Higher derivatives control the approximation accuracy.

Example

Give a bound on the error, when e^x is approximated by its fourth-degree Taylor polynomial about 0 for $-0.5 \leq x \leq 0.5$.

Multivariate Taylor Polynomials

Let $f : \mathbb{R}^n \to \mathbb{R}$ be smooth near $\mathbf{a} = (a_1, \ldots, a_n)$.

First-Order Approximation (Linearization)

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$$

where $\nabla f(\mathbf{a}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{a}) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{a}) \end{bmatrix}$.

Second-Order Approximation

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^\top H_f(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

where $H_f(\mathbf{a})$ is the Hessian matrix of second derivatives.

Higher-Order Multivariate Taylor Polynomials

Multi-index Notation

For $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$:

$$|\alpha| = \alpha_1 + \cdots + \alpha_n, \quad \alpha! = \alpha_1! \cdots \alpha_n!, \quad (\mathbf{x} - \mathbf{a})^\alpha = \prod_{i=1}^{n} (x_i - a_i)^{\alpha_i}.$$

Taylor Polynomial of Order m

$$T_m(\mathbf{x}) = \sum_{|\alpha| \le m} \frac{D^\alpha f(\mathbf{a})}{\alpha!} (\mathbf{x} - \mathbf{a})^\alpha$$

where $D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$.

- The terms of degree 1 and 2 give the linear and quadratic approximations.
- Higher-degree terms capture curvature and mixed derivatives.

Mathematics Beyond Calculus for Machine Learning

Mathematics Beyond Calculus for Machine Learning

- Calculus is only the beginning: it gives local, continuous reasoning.
- Modern ML models rely on more advanced tools:
    - Probability and measure theory
    - Functional analysis
    - Differential geometry
    - Differential equations and dynamics
- These fields provide both rigor and intuition for high-dimensional learning.

Probability: The Powerhouse of Machine Learning

- Data is uncertain $\rightarrow$ need probabilistic reasoning.
- Key ideas:
  - Bayesian inference: learning as belief updating.
  - Stochastic optimization (SGD): random sampling for efficiency.
  - Generative models: learning probability distributions.
- Measure-theoretic probability underpins likelihoods, expectations, and convergence proofs.

Differential Equations and Dynamics

- Neural network training can be viewed as a dynamical system:

$$\frac{d\theta}{dt} = -\nabla_\theta L(\theta)$$

- Continuous-time limits $\rightarrow$ differential equations:
  - Gradient flow, diffusion processes.
  - Neural ODEs model data with continuous transformations.
- Understanding stability and convergence uses ODE theory.

Differential Geometry and Computer Vision

- Data often lives on a **manifold**, not in flat space.
- Geometry provides tools for:
  - Manifold learning, embeddings (diffusion maps, Laplacian eigenmaps)
  - Invariance to rotations, translations $\rightarrow$ Lie groups
  - Shape analysis and 3D perception
- Vision models (e.g., equivariant CNNs) are grounded in geometric principles.

Functional Analysis and Infinite-Dimensional Spaces

- Functions can be seen as points in a vector space.
- Hilbert and Banach spaces generalize geometry to function spaces.
- Reproducing Kernel Hilbert Spaces (RKHS) form the backbone of kernel methods and Gaussian processes.