10-701 Recitation 1

Linear Algebra Review

Jin Sun

Administrative stuff

- Course website: http://alex.smola.org/teaching/10-701-15/
- Autolab: https://autolab.cs.cmu.edu/
- Piazza: https://piazza.com/class/i4ivtbjbrt219e
- Theoretical Assignments: submit **pdf** files (*.pdf)
  - Use provided latex source file
  - MS word or other text editors, clearly mark your problems
  - Scan handwriting sheets, make sure we can recognize your handwriting
- Programming Assignments: submit code.tar, compressed from "code" folder in the handout folder
  - Not handout.tar, do not submit extra files
  - Unlimited submission
  - More information on Piazza

More administrative stuff

- Recitation: Thursday 4-5pm HH B131
  - Slides and videos will be posted
- TA office hours (for all TAs): Thursday 5-6pm after recitation or by appointment
- We do not debug for students

Our team

- Instructor: Alex Smola
- TAs and tasks in charge (in general):
  - Jay-Yoon Lee: Homework
  - Jin Sun: Autolab (programming assignments)
  - Shen Wu: Piazza
  - Di Xu: Project
  - Zhou Yu: Recitations

Nice materials

- Linear Algebra Review from Zico Kolter
  - http://www.cs.cmu.edu/~zkolter/course/linalg/index.html
- Linear Algebra Review from Jing Xiang
  - http://www.cs.cmu.edu/~jingx/docs/linearalgebra.pdf
- The Matrix Cookbook
  - http://www.mit.edu/~wingated/stuff_i_use/matrix_cookbook.pdf
- Probability Review from Aaditya Ramdas
  - http://www.cs.cmu.edu/~aramdas/videos.html

Linear algebra review

- Basics
- Property of Matrices
- Vector Norms
- Matrix Calculus
- An example: Linear Regression
- Eigen Decomposition
- Quadratic Form
- Singular Value Decomposition

* Many slides are from Jing Xiang's linear algebra review sheet

Basics

- Vectors and matrices
  - Vector $\boldsymbol{x} = (x_1, x_2, \dots, x_n) \in \mathcal{R}^n$
    - Implicitly means column vector
  - Matrix $\boldsymbol{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \cdots & \cdots & \cdots \\ x_{m,1} & \cdots & x_{m,n} \end{bmatrix} \in \mathcal{R}^{m \times n}$

Vector product

- Vector Product
  - Inner product (dot product): Result is a scalar
    - $\boldsymbol{u} = (u_1, u_2, \dots, u_n) \in \mathcal{R}^n$, $\boldsymbol{v} = (v_1, v_2, \dots, v_n) \in \mathcal{R}^n$, column vectors
    - $< \boldsymbol{u}, \boldsymbol{v} > = \sum_{i=1}^{n} u_i v_i$
    - Other forms: $\boldsymbol{u}^T \boldsymbol{v}$
    - A measurement for similarity
  - Outer product (cross product): Result is a matrix $\mathcal{R}^{m \times n}$
    - $\boldsymbol{u} = (u_1, u_2, \dots, u_m) \in \mathcal{R}^m$, $\boldsymbol{v} = (v_1, v_2, \dots, v_n) \in \mathcal{R}^n$, column vector
    - $\boldsymbol{u} \otimes \boldsymbol{v} = \begin{bmatrix} u_1 v_1 & \dots & u_1 v_n \\ \dots & \dots & \dots \\ u_m v_1 & \dots & u_m v_n \end{bmatrix}$
    - Other forms: $\boldsymbol{u}\boldsymbol{v}^{\boldsymbol{T}}$

Matrix multiplication

- Matrix multiplication
  - If $A \in \mathcal{R}^{m\times n}, B \in \mathcal{R}^{p\times q}$, AB is defined only when $n = p$, the result is $\mathcal{R}^{m\times q}$
  - Associative: $(AB)C = A(BC)$
  - Distributive: $A(B + C) = AB + AC$
  - **NOT** commutative: $AB \neq BA$, may not even be defined

Matrix multiplication as vector product

Inner product

- $A \in \mathcal{R}^{m \times n}, B \in \mathcal{R}^{n \times p}$
- $\boldsymbol{a}_i \in \mathcal{R}^{1 \times n}$ is a row of A, and $\boldsymbol{b}_j \in \mathcal{R}^{n}$ is a column of B
- $AB = \begin{bmatrix} \boldsymbol{a}_1\boldsymbol{b}_1 & \dots & \boldsymbol{a}_1\boldsymbol{b}_p \\ \dots & \dots & \dots \\ \boldsymbol{a}_m\boldsymbol{b}_1 & \dots & \boldsymbol{a}_m\boldsymbol{b}_p \end{bmatrix}$

Outer product

- $A \in \mathcal{R}^{m \times n}, B \in \mathcal{R}^{n \times p}$
- $\boldsymbol{a}_i \in \mathcal{R}^{m}$ is a column of A, and $\boldsymbol{b}_j \in \mathcal{R}^{1 \times p}$ is a row of B
- $AB = \sum_{i,j} \boldsymbol{a}_i \boldsymbol{b}_j$

Transpose

- $A \in \mathcal{R}^{m \times n}$, $A^T \in \mathcal{R}^{m \times n}$
- $A_{i,j} = A_{j,i}^T$
- $(A^T)^T = A$
- $(AB)^T = B^T A^T$
- $(A + B)^T = (B + A)^T$

Rank

- A set of vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ is linear independent if not one of them can be represented as a linear combination of the rest
- $Rank(A)$ is the size of the largest collection of linearly independent columns (or rows) of A. In fact, column rank is equal to row rank.
- $A \in \mathcal{R}^{m \times n}$ is full rank if $Rank(A) = \min(m, n)$, otherwise it is low rank
- $Rank(A^T) = Rank(A)$

Inverse

- A matrix is invertible only if
  - it is square
  - it is full rank (or many other equivalent conditions, we'll see later)

- If $A \in \mathcal{R}^{n \times n}$, $A^{-1} \in \mathcal{R}^{n \times n}$
- $A^{-1}A = AA^{-1}$
- $(A^{-1})^{-1} = A$
- $(A^{-1})^{T} = (A^{T})^{-1}$
- $(AB)^{-1} = B^{-1}A^{-1}$

Trace

- The trace of a square matrix is the sum of its diagonal elements
- $Tr(A) = \sum_{i=1}^{n} A_{ii}\,, A \in \mathcal{R}^{n\times n}$

For $A, B \in \mathcal{R}^{n\times n}$

- $Tr(A^T B) = Tr(B^T A) = Tr(AB^T) = Tr(BA^T) = \sum_{i,j}^{n} A_{i,j} B_{i,j}$
- $Tr(A) = Tr(A^T)$
- $Tr(A + B) = Tr(B + A)$
- $Tr(cA) = c\,Tr(A)$

Vector norms

Norm – a measurement of magnitude

- Family of norms: $||\boldsymbol{x}||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$
- l_1 norm: $||\boldsymbol{x}||_1 = \sum_{i=1}^{n} |x_i|$
- l_2 norm: $||\boldsymbol{x}||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$ --- Euclidean distance
- l_∞ norm: $||\boldsymbol{x}||_\infty = \max(|x_i|)$
- l_0 norm: $||\boldsymbol{x}||_0 = \#(x_i \neq 0)$

Matrix calculus

- Denominator layout
- Gradient:
  - If $f: \mathcal{R}^n \to \mathcal{R}$, $\nabla f \in \mathcal{R}^n$,$\nabla f_i = \frac{\partial f}{\partial x_i}$
  - If $f: \mathcal{R}^{m\times n} \to \mathcal{R}$, $\nabla f \in \mathcal{R}^{m\times n}$, $\nabla f_{i,j} = \frac{\partial f}{\partial x_{i,j}}$
  - If $f: \mathcal{R}^n \to \mathcal{R}^m$, $\nabla f \in \mathcal{R}^{m\times n}$, $\nabla f_{i,j} = \frac{\partial f_i}{\partial x_j}$
- Hessian:
  - If $f: \mathcal{R}^n \to \mathcal{R}$, $\nabla^2 f \in \mathcal{R}^n$, $\nabla^2 f_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$

Matrix calculus

- Chain Rule: (credit to Bhiksha Raj's MLSP class)
  - If $y = f_1(f_2(f_3(\dots f_k(\boldsymbol{X}))))$ is a composition of functions.
  - $\frac{dy}{d\boldsymbol{X}} = \left(\frac{df_k}{d\boldsymbol{X}}\right)^T \left(\frac{df_{k-1}}{df_k}\right)^T \left(\frac{df_{k-2}}{df_{k-1}}\right)^T \dots \left(\frac{df_2}{df_3}\right)^T \frac{df_1}{df_2}$
- Useful derivatives: (look at Jing's review and matrix cookbook)
  - $\frac{\partial(\boldsymbol{a}^T\boldsymbol{x})}{\partial\boldsymbol{x}} = \frac{\partial(\boldsymbol{x}^T\boldsymbol{a})}{\partial\boldsymbol{x}} = \boldsymbol{a}, \frac{\partial(\boldsymbol{x}^T\boldsymbol{A})}{\partial\boldsymbol{x}} = \boldsymbol{A}$
  - $\frac{\partial(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}} = (\boldsymbol{A} + \boldsymbol{A}^T)\boldsymbol{x}$
  - $\frac{\partial\boldsymbol{x}^T}{\partial\boldsymbol{x}} = \boldsymbol{I}$
  - …

Linear regression

- Work out the normal equation:
- Objective: $\underset{w}{\text{minimize}} \frac{1}{2}||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}||_2^2$
  - where $\boldsymbol{y} \in \mathcal{R}^n$, $\boldsymbol{X} \in \mathcal{R}^{n \times f}$, $\boldsymbol{w} \in \mathcal{R}^f$

Solution

- Expand the expression
  - $f = \frac{1}{2}||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}||_2^2 = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}) = \frac{1}{2}(\boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{w}^T\boldsymbol{X}^T\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{X}\boldsymbol{w} + \boldsymbol{w}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w}) = \frac{1}{2}(\boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{w}^T\boldsymbol{X}^T\boldsymbol{y} + \boldsymbol{w}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w})$
- Take derivative
  - $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{w}} = -\boldsymbol{X}^T\boldsymbol{y} + \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w}$
  - Solve it using chain rule?
- Set it to zero
  - $\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^T\boldsymbol{y}$
  - Why is this ill conditioned?

Eigen decomposition

- $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}$, $\boldsymbol{A} \in \mathcal{R}^{n \times n}$
- Each column of Q is an Eigen vector. Λ is a diagonal matrix with each element as a Eigen value.
- $\boldsymbol{Au} = \lambda\boldsymbol{u}$, for an Eigen vector $\boldsymbol{u}$ and its Eigen vector λ.
- Eigen vectors have unit length and are orthogonal to each other.
- Zero Eigen values indicate low rank.
- Relation to Principle Component Analysis.
- $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{T}$, when $\boldsymbol{A}$ is symmetric

Quadratic form

- Definiteness
  - $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \sum_{\boldsymbol{i},\boldsymbol{j}} \boldsymbol{A}_{\boldsymbol{i},\boldsymbol{j}}\boldsymbol{x}_{\boldsymbol{i}}\boldsymbol{x}_{\boldsymbol{j}}$, $\boldsymbol{A} \in \mathcal{R}^{n\times n}$
  - Positive definite, $\boldsymbol{A} \succ 0$: $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$, for all non-zero $\boldsymbol{x}$
  - Semi-positive definite, $\boldsymbol{A} \geq 0$: $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geq 0$, for all non-zero $\boldsymbol{x}$
- $\boldsymbol{A} \succ 0$: All Eigen values are positive→ full rank →invertible
- $\boldsymbol{A} \geq 0$: All Eigen values are non-negative.

- Covariance matrix is always positive-semi definite
  - $\boldsymbol{x}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{x} = ||\boldsymbol{B}\boldsymbol{x}||_2^2 \geq 0 \rightarrow \boldsymbol{B}^T\boldsymbol{B} \geq \boldsymbol{0}$

Singular value decomposition

- $\boldsymbol{A} = \boldsymbol{U\Sigma V}^T, \boldsymbol{A} \in \mathcal{R}^{m\times n}, \boldsymbol{U} \in \mathcal{R}^{m\times m}, \boldsymbol{\Sigma} \in \mathcal{R}^{m\times n}, \boldsymbol{V} \in \mathcal{R}^{n\times n}$
- $\boldsymbol{\Sigma}$ is a (rectangle) diagonal matrix with singular values.
- $\boldsymbol{U}$ and $\boldsymbol{V}$ are matrices containing left and right singular vectors (orthogonal basis).
- Think it as Eigen decomposition:
  - $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{V\Sigma^T U^T U\Sigma V}^T = \boldsymbol{V\Sigma^T\Sigma V^T} = \boldsymbol{VPV}^T$
  - $\boldsymbol{AA}^T = \boldsymbol{U\Sigma V}^T\boldsymbol{V\Sigma^T U^T} = \boldsymbol{U\Sigma\Sigma^T U^T} = \boldsymbol{UQU}^T$
  - $\boldsymbol{A}^T\boldsymbol{A}$ and $\boldsymbol{AA}^T$ are symmetric matrices.