Weekly Report(Apr.30,2018-May.13,2018)

Zhang Yuandi

Abstract

Sorry for my delayed submit of Weekly Report due to midterm examination review. In the last two weeks, I have finished the last three weeks of **Linear Algebra** and learned the first two units of **Introduction to Probability**.

1 Work done in these weeks

1.1 Linear Algebra

The last courses focus on Orthogonality and Eigenvalues.

1.1.1 Orthogonality

Orthogonal Vectors

Vectors x and y are considered to be orthogonal (Perpendicular) if they meet at a right angle. Using the Euclidean length

$$\|\mathrm{x}\|_2 = \sqrt{x^2{}_0 + x^2{}_1 + \ldots + x^2{}_{n-1}} = \sqrt{x^T x}$$

we find that the Pythagorean Theorem dictates that if the angle in the triangle where x and y meet is a right angle, then $\|z\|_2{}^2 = \|x\|_2{}^2 + \|y\|_2{}^2$. In this case, we can get that $x^T y = 0$.

Orthogonal Spaces

Let **V, W** $\subset \mathrm{R}^n$ be subspaces. Then **V** and **W** are said to be orthogonal if and only if $v \in$ **V** and $w \in$ **W** implies that

$$v^T w = 0$$

.We will use the notation **V** $\perp$ **W** to indicate that subspace **V** is orthogonal to subspace **W**.In other words: Two subspaces are orthogonal if all the vectors from one of the subspaces are orthogonal to all of the vectors from the other subspace.

Solving the Normal Equations

We have solved the normal equations

$$A^T Ax = A^T b,$$

where A $\in \mathrm{R}^{m*n}$ has linear independent columns, via the following steps:

- Form $y = A^T b$.
- Form $A^T A$.
- Invert $A^T A to compute B = (A^T A)^{-1}$.
- Compute $\hat{x} = By = (A^T A)^{-1} A^T b$.

Then, let's focus on how to use the Cholesky factorization. Here are the steps:

- Compute $C = A^T A$.
- Compute the Cholesky factorization $C = LL^T$, where L is lower triangular.
- Compute the Cholesky factorization $C = LL^T$, where L is lower triangular.This allows us to ake advantage of symmetry in C.
- Compute $y = A^T b$.
- Solve $Lz = y$.
- Solve $L^T \hat{x} = z$.

The vector $\hat{x}$ is then the best solution (in the linear least-squares sense) to $Ax \approx b$.That is, a symmetric matrix $C \in R^{m*n}$ is said to be symmetric positive definite if $x^T C x \geq 0$ for all nonzero vectors $x \in R^m$.

Orthonormal Bases Let $q_0, q_1, ..., q_{k-1} \in R^m$. Then these vectors are (mutually) orthonormal id for all $0 \leq i, j < k$

$$q_i^T q_j = \begin{cases} 1 & \text{if } i = j \\ 0 & otherwise. \end{cases}$$

And how to get orthogonal bases? We can get them by **Gram-Schmidt orthogonalization**.

$$a_k^\perp = a_k - q_0^T a_k q_0 - q_1^T a_k q_1 - ...q_k - 1^T a_k q_{k-1}.$$

Take $a_k^\perp$, the component of a_k orthogonal to $q_0, q_1, ...q_{k-1}$, and make it of unit length:

$$q_k = a_k^\perp / \parallel a_k^\perp \parallel_2,$$

And we will see that Span($a_0, a_1, ..., a_k$) = Span($q_0, q_1, ..., q_{k-1}$).

1.1.2 Eigenvalues

The Algebraic Eigenvalue Problem The algebraic eigenvalue problem is given by

$$Ax = \lambda x.$$

And here are some equivalent statements:

- Ax = λ x, where x $\neq$ 0.
- Ax - λ x = 0, where x $\neq$ 0.
- Ax - λ Ix = 0, where x $\neq$ 0.
- (A - λ I)x = 0, where x $\neq$ 0.
- A - λ I is singular.
- N(A - λ I) contains a nonzero vector x.
- dim(N(A - λ I)) $>$ 0.

If we find a vector $\neq$ 0 such that $Ax = \lambda x$, it is certainly not unique.

- For any scalar α, $A(\alpha x) = \lambda(\alpha x)$ also holds.
- If $Ax = \lambda x$ and $Ay = \lambda y$, then $A(x + y) = Ax + Ay = \lambda x + \lambda y = \lambda(x + y)$.

We conclude that the set of all vectors x that vectors x that satisfy $Ax = \lambda x$ is a subspace.

Practical Methods for Computing Eigenvectors and Eigenvalues

Let us assume that $A \in \mathbb{R}^{m*n}$ is diagonalizable so that

$$A = V \Lambda V^{-1} = (\upsilon_0 \ \upsilon_1 \ \cdots \ \upsilon_{n-2} \ \upsilon_{n-1}) \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-2} & 0 \\ 0 & 0 & \cdots & 0 & \lambda_{n-1} \end{pmatrix} (\upsilon_0 \ \upsilon_1 \ \cdots \ \upsilon_{n-2} \ \upsilon_{n-1})^{-1}$$

This means that υ_i is an eigenvector associated with λ_i.Then, we start with some vector $x^{(0)} \in \mathbb{R}^n$. Since **V** is nonsingular, the vectors $\upsilon_0, \ldots, \upsilon_{n-1}$ form a linearly independent bases for $\mathbb{R}^n$. Hence,

$$x^{(0)} = \gamma_0\upsilon_0+\gamma_1\upsilon_1+\cdots+\gamma_{n-2}\upsilon_{n-2}+\gamma_{n-1}\nu_{n-1} = (\nu_0 \quad \nu_1 \quad \cdots \quad \nu_{n-2} \quad \nu_{n-1}) \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{n-2} \\ \gamma_{n-1} \end{pmatrix} = Vc.$$

Now, we generate

$$x^{(1)} = A^{-1}x^{(0)}$$
$$x^{(2)} = A^{-1}x^{(1)}$$
$$x^{(3)} = A^{-1}x^{(2)}$$
$$\vdots$$

Then, we can get

$$x^{(k)} = A^{-1}x^{(k-1)} = (A^{-1})^2x^{(k-2)} = \cdots = (A^{-1})^kx^{(0)}.$$
$$x^{(1)} = \lambda_{n-1}A^{-1}x^{(0)}$$
$$x^{(2)} = \lambda_{n-1}A^{-1}x^{(1)}$$
$$x^{(3)} = \lambda_{n-1}A^{-1}x^{(2)}$$
$$\vdots$$

Then,

$$x^{(k)} = \gamma_0|\frac{\lambda_{n-1}-\mu}{\lambda_0-\mu}|^k\nu_0 + \gamma_1|\frac{\lambda_{n-1}-\mu}{\lambda_1-\mu}|^k\nu_1 + \cdots + \gamma_{n-2}|\frac{\lambda_{n-1}-\mu}{\lambda_{n-2}-\mu}|^k\nu_{n-2} + \gamma_{n-1}\nu_{n-1}$$

If we knew ν_{n-1} but not λ_{n-1}, then we could compute the Rayleigh quotient:

$$\lambda_{n-1} = \frac{\nu_{n-1}^T A\nu_{n-1}}{\nu_{n-1}^T\nu_{n-1}}.$$

And by using the approximation we can pick

$$\mu = \frac{x^{(k)T}Ax^{(k)}}{x^{(k)T}x^{(k)}} \approx \lambda_{n-1}$$

1.2 Probability

The world is filled with uncertainty, and probability is a part of scientific literacy. With probability, we are able to fight with randomness as much as we can. The first two units focus on some basic models and axioms.

1.2.1 Sets

Sets is a collection of distinct elements. If can contain both finite and infinite elements. We can get the unions or intersections of different sets.

1.2.2 Probability Models and Axioms

- The **sample space** Ω is the set of all possible outcomes of an experiment.
- The **probability law**, which assigns to a set A of possible outcomes (also called an event) a nonnegative number P(A)(called the probability of A).
- P(A) $\geq$ 0, for every event A.

- If the sample space has an infinite number od elements and A_1, A_2, ... is a sequence of disjoint events, then the probability of there union satisfies
  $P(A_1 \bigcup A_2 \bigcup \cdots) = P(A_1) + P(A_2) + \cdots$.
- $P(\Omega) = 1$.
- The probability of any event $s_1, s_2, \ldots, s_n$ is the sum of the probability of its elements:

$$P(s_1, s_2, \ldots, s_n) = P(s_1) + P(s_2) + \cdots + P(s_n).$$

- If the sample space consists of n possible outcomes which are equally likely, then

$$P(A) = \frac{number\ \ of\ \ elements\ \ of\ \ A}{n}.$$

1.2.3 Some Properties of Probability Laws

Let A, B and C be events.

(a) If $A \subset B$, then $P(A) \leq P(B)$.

(b) $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$.

(c) $P(A \bigcup B) \leq P(A) + P(B)$.

(d) $P(A \bigcup B \bigcup C) = P(A) + P(A^c \bigcap b) + P(A^c \bigcap B^c \bigcap C)$.

1.3 LaTeX

Thanks for the handbook my buddy recommended, it is much more convenient and interesting for me to use LaTeXand it helps me understand how LaTeXworks better. It really helps me a lot and I'm very appreciated to it. I have looked through some pages of the handbook but not finished.

2 Questions for My Studying Pace

I planned to learn every basic course the training pdf mentions, but it seems to be quite a lot and some courses I have learned already in SJTU. And I'm running out of time for this semester is reaching its ending. I'm afraid that I don't have enough time to finish the training. How should I adjust my studying pace?

3 Plan for the Next Weeks

1. Learn the unit2, unit3, unit4, unit5, unit6 courses of **Probability**.
2. Learn the course **How to Use Git and GitHub**