Mathematics Essential*

Ke Chen

To carry out this non-assessed coursework, you can use the Jupyter notebook, `math.ipynb`, which facilitates your doing the practical part of this coursework and allows you to use markdown cells to answer Math questions. The file is available on BlackBoard alongside this document.

In essence, mathematics lays the foundation for machine learning including representation learning. Thus, it is necessary for one who wants to understand the principle of machine learning and probes into machine learning algorithms. In fact, a machine learning practitioner who masters such transferrable knowledge may significantly distinguish from those who simply treat machine learning algorithms "blackbox" and is hence more likely to achieve a success in real applications thanks to "know-how". In this non-assessed coursework, you are asked to do several math exercises which involves the math knowledge not only immediately required by lectures and courseworks in this course unit but also transferrable to other machine learning areas. The exercises consist of two parts: practicals with programming in Python and math questions.

1 Practicals

In this work, you are asked to make use of built-in functions in Python to learn and understand some linear algebra essentials required by coursework and lectures. In this course unit (and most of machine learning textbooks), the **column vector** notation is adopted; i.e., in a data matrix, each column corresponds to a data entity (point). For instance, $X_{d\times|X|} = \{X_{ij}\}_{d\times|X|} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{|X|}\}$ means that there are $|X|$ data points in X and each data point, $\boldsymbol{x}_i$, has d features or is a d-dimensional feature vector for $i = 1, 2, \cdots, |X|$. In Python, however, built-in functions in most libraries use the **row vector** notation. So you should pay attention to this difference and check if you need to transpose a data matrix whenever using built-in functions in doing your coursework.

In your answer notebook, `math.ipynb`, you are given with two matrices, $A_{5\times3}$ and $B_{3\times5}$. By using two matrices, A and B, do the following assignments[1]:

Assignment 1. Apply the built-in function, `numpy.cov`, in the `numpy` library to A and B to get their covariance matrices, $\mathrm{cov}(A)$ and $\mathrm{cov}(B)$.

Assignment 2. Apply the built-in function, `np.linalg.eig`, in the `numpy` library to $\mathrm{cov}(A)$ and $\mathrm{cov}(B)$ to achieve all the eigenvalues and their corresponding eigenvectors, respectively.

Assignment 3. Calculate $X = V\mathrm{diag}(\boldsymbol{\lambda})V^{-1}$ and report your observation by comparing two matrices: X and $\mathrm{cov}(B)$, where V is the matrix constructed by eigenvectors of $\mathrm{cov}(B)$ and $\mathrm{diag}(\boldsymbol{\lambda})$ is

*****Non-Assessed Coursework**: you are expected to complete this coursework within one week.

[1]For the definition and concepts involved in the assignements, see the "Mathematics Essentials" lecture note.

the diagonal matrix consisting of the corresponding eigenvalues of $\mathrm{cov}(B)$ you have achieved in **Assignment 2**.

Assignment 4. Apply the built-in function, `np.linalg.svd`, in the `numpy` library to A and B to get the SVD matrix decomposition, respectively.

In a high-dimensional space, there often exist surprising facts beyond our observations in 3-D real world. You are asked to replicate the result of a surprising fact presented in the lecture with your own code in Python.

Assignment 5. Let V_{hc} and V_{hs} denote volumes of a unit hyper-cube and its embedding hyper-sphere, respectively[2]. In your answer notebook, calculate volumes of the unit hyper-cube and its embedding hyper-sphere and draw a diagram to visualise the ratios between V_{hs} and V_{hc} (V_{hs}/V_{hc}) up to 20 dimensions. (**Hint**: you can use the built-in function, `scipy.special.gamma`, in the `scipy` library for Gamma function computation.)

2 Mathematics

In this work, you are asked to make use of your Math knowledge and literature review skills acquired in your UG study along with the reinforcement given in the Math lecture to do several theoretical assignments of which solutions will be used in subsequent lectures.

Assignment 6. Making use of on-line resources, e.g. Wikipedia, understand the following algebraic/math concepts and operations along with their **geometric meaning** (if there is): (a) inner product between two vectors; (b) outer product between two vectors; (c) determinant of a square matrix; (d) trace of a square matrix; (e) positive definite and semi-definite square matrix; and (f) graph theory: directed vs. undirected graph, weighted vs. unweighted graph, and shortest path.

Assignment 7. Given a data matrix, $X_{d\times|X|} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{|X|}\}$, where $\boldsymbol{x}_i$ is a d-dimensional vector corresponding to the i-th data entity (point) in the dataset, its covariance matrix, S_X, in the element-wise form is defined in the lecture note[3].

(a) Show this covariance matrix, S_X, can be expressed in the following vectorial form:

$$S_X = \frac{1}{|X|-1}\sum_{i=1}^{|X|}\Big(\boldsymbol{x}_i - \boldsymbol{m}\Big)\Big(\boldsymbol{x}_i - \boldsymbol{m}\Big)^T, \quad \text{where } \boldsymbol{m} = \frac{1}{|X|}\sum_{i=1}^{|X|}\boldsymbol{x}_i.$$

(b) Let $\hat{X}_{d\times|X|} = \{\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \cdots, \hat{\boldsymbol{x}}_{|X|}\}$ be the centralised matrix of X, where $\hat{\boldsymbol{x}}_i = \boldsymbol{x}_i - \boldsymbol{m}$. Furthermore, show this covariance matrix, S_X, is also expressed in the following matrix form:

$$S_x = \frac{1}{|X|-1}\hat{X}\hat{X}^T.$$

[2] Formulas of hyper-cube and hyper-sphere volumes, see the “Representation Learning Overview” lecture note.

[3] For the element-wise covariance matrix definition, see the “Mathematics Essentials” lecture note.

Hints: (1) You can expand the vectorial or matrix form of a covariance matrix to be its element-wise form as defined in the lecture note. (2) You may start with a special case when $d = 2$ then extend it to any dimension (this hint is also applicable to the next assignment).

Assignment 8. Given a function, $f(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{x}$, where $\boldsymbol{x}$ is a d-dimensional vector and A is a $d \times d$ constant symmetric matrix, prove

$$\frac{df(\boldsymbol{x})}{d\boldsymbol{x}} = 2A\boldsymbol{x} - 2\boldsymbol{x} = 2(A - I_d)\boldsymbol{x},$$

where I_d is the $d \times d$ identity matrix.

Assignment 9. Given a composite function, $y(\Theta) = g\big(\omega f(\boldsymbol{w}^T\boldsymbol{x} + u) + v\big)$, where $\Theta = (\omega, \boldsymbol{w}, u, v)$ is a collective nation of all the variables and $\boldsymbol{x}$ is a constant vector. $f(\cdot)$ and $g(\cdot)$ are differentiable functions with the gradients, $f'(\cdot)$ and $g'(\cdot)$, respectively, prove

$$\frac{\partial y(\Theta)}{\partial \omega} = g'\big(\omega f(\boldsymbol{w}^T\boldsymbol{x} + u) + v\big) f(\boldsymbol{w}^T\boldsymbol{x} + u),$$

$$\frac{\partial y(\Theta)}{\partial v} = g'\big(\omega f(\boldsymbol{w}^T\boldsymbol{x} + u) + v\big),$$

$$\frac{\partial y(\Theta)}{\partial \boldsymbol{w}} = \omega g'\big(\omega f(\boldsymbol{w}^T\boldsymbol{x} + u) + v\big) f'(\boldsymbol{w}^T\boldsymbol{x} + u)\boldsymbol{x},$$

and

$$\frac{\partial y(\Theta)}{\partial u} = \omega g'\big(\omega f(\boldsymbol{w}^T\boldsymbol{x} + u) + v\big) f'(\boldsymbol{w}^T\boldsymbol{x} + u).$$

Hint: You need to apply the derivative chain rule in your proof.