Answers Machine Learning Exercises 1 - Corrected Version

Tim van Erven

October 1, 2007

1 Distributivity of Matrix Multiplication

Suppose A and B are arbitrary $k \times m$ matrices and C and D are arbitrary $m \times n$ matrices. Recall that the $k \times n$ matrix product AC is defined component-wise as:

$$(AC)_{ij} = \langle \mathbf{a}_i, \mathbf{c}_j \rangle = \sum_{l=1}^{m} A_{il} C_{lj} \quad \text{for } i = 1, \ldots, k, j = 1, \ldots, n,$$

where $\mathbf{a}_i^\top$ is the ith row vector of matrix A and $\mathbf{c}_j$ is the jth column vector of matrix C.

Exercise 1. *Show that matrix multiplication is distributive.* That means you have to show that

1. $(A + B)C = AC + BC$, and
2. $A(C + D) = AC + AD$.

Hint 1: Both cases can be proved in very similar ways. Your proof of the first case might have the following outline:

1. Show that the matrix that is the result of $(A + B)C$ has the same dimensions as the matrix that is the result of $AC + BC$. In other words, if $(A+B)C$ is an $l \times p$ matrix, then $AC+BC$ should also be an $l \times p$ matrix.

2. Now you need to show that all components of $(A + B)C$ and $AC + BC$ are identical. This can be done by proving that the matrices are identical in an arbitrary component:

$$((A + B)C)_{ij} = (AC + BC)_{ij} \quad \text{for arbitrary } i \text{ and } j.$$

   This can be done by writing out both sides of the equation using their definitions and then doing some algebraic manipulations.

Hint 2: Try with vectors instead of matrices first if you're having difficulties or work out an example with numbers.

Answer Exercise 1. The following proof is very long. Don't get intimidated by that. It is long because it is very detailed. In your own proofs it would be preferable to provide fewer details to make the proof shorter.

I will follow the outline of the proof given in Hint 1.

Part 1 We first show that $(A+B)C$ and $AC+BC$ have the same dimensionality: As A and B are both $k \times m$ matrices, so is their sum $A+B$. Together with the fact that C is an $m \times n$ matrix this implies that the matrix product $(A+B)C$ is a $k \times n$ matrix. We can also see directly that both AC and BC are $k \times n$ matrices, which implies that their sum $AC+BC$ has the same dimensions. This verifies that both $(A+B)C$ and $AC+BC$ are $k \times n$ matrices.

By a similar argument it can be shown that $A(C+D)$ and $AC+AD$ are both $k \times n$ matrices.

Part 2 Now we will show that all components of $(A+B)C$ and $AC+BC$ are identical. For this it is sufficient to show that

$$((A+B)C)_{ij} = (AC+BC)_{ij} \tag{1}$$

for any arbitrary row i and column j.

By definition of matrix addition, it follows that the ith row vector of the matrix $A+B$ is equal to $\mathbf{a}_i^\top + \mathbf{b}_i^\top$, where $\mathbf{a}_i^\top$ and $\mathbf{b}_i^\top$ denote the ith row vectors of A and B, respectively. Let $\mathbf{c}_j$ denote the jth column vector of C. We now have by definition of matrix multiplication that

$$((A+B)C)_{ij} = \langle \mathbf{a}_i + \mathbf{b}_i, \mathbf{c}_j \rangle \tag{2}$$

and

$$(AC+BC)_{ij} = (AC)_{ij} + (BC)_{ij} = \langle \mathbf{a}_i, \mathbf{c}_j \rangle + \langle \mathbf{b}_i, \mathbf{c}_j \rangle. \tag{3}$$

Recall that $\mathbf{a}_i^\top$ and $\mathbf{b}_i^\top$ are row vectors from the matrices A and B, and that $\mathbf{c}_j$ is a column vector from the matrix C. It follows by writing out the inner products in Equations 2 and 3 using the definition of the inner product, that

$$\begin{aligned}
\langle \mathbf{a}_i + \mathbf{b}_i, \mathbf{c}_j \rangle &= \sum_{l=1}^{m} (A_{il} + B_{il}) \cdot C_{lj} \\
&= \sum_{l=1}^{m} (A_{il} \cdot C_{lj} + B_{il} \cdot C_{lj}) \\
&= \sum_{l=1}^{m} A_{il} \cdot C_{lj} + \sum_{l=1}^{m} B_{il} \cdot C_{lj} \\
&= \langle \mathbf{a}_i, \mathbf{c}_j \rangle + \langle \mathbf{b}_i, \mathbf{c}_j \rangle. \qquad (4)
\end{aligned}$$

This may look like a pretty long derivation, but notice that the first step follows by writing out the definition of the inner product $\langle \mathbf{a}_i + \mathbf{b}_i, \mathbf{c}_j \rangle$ and the last step follows by writing out the definitions of the inner products $\langle \mathbf{a}_i, \mathbf{c}_j \rangle$ and $\langle \mathbf{b}_i, \mathbf{c}_j \rangle$, so there is really only one step in between. Together with Equations 2 and 3 this proves Equation 1, which was what we were after.

Let $\mathbf{d}_j$ denote the jth column vector of D. By a similar argument as above we get for arbitrary row i and column j that

$$\begin{aligned}
(A(C+D))_{ij} &= \langle \mathbf{a}_i, \mathbf{c}_j + \mathbf{d}_j \rangle \\
&= \sum_{l=1}^{m} A_{il} \cdot (C_{lj} + D_{lj}) \\
&= \sum_{l=1}^{m} (A_{il} \cdot C_{lj} + A_{il} \cdot D_{lj}) \\
&= \sum_{l=1}^{m} A_{il} \cdot C_{lj} + \sum_{l=1}^{m} A_{il} \cdot D_{lj} \\
&= \langle \mathbf{a}_i, \mathbf{c}_j \rangle + \langle \mathbf{a}_i, \mathbf{d}_j \rangle \\
&= (AC)_{ij} + (AD)_{ij} \\
&= (AC + AD)_{ij}, \qquad (5)
\end{aligned}$$

which completes the proof.

Grading of Exercise 1:

- Part 1 of the proof is perhaps nitpicking, so you still get full marks if you have skipped that step.
- You can get at most 5 points for Exercise 1. Partial points are awarded for any intermediate steps of the proof that can be identified.
- Because the two cases ($(A+B)C = AC + BC$ and $A(C+D) = AC+AD$) are so similar, you get points if you get either one of them correctly.
- In particular you will get
  - 1 point for either
    - showing Part 1 or
    - not doing Part 1 at all and getting *both* of the following two definitions correct,
  - 1 point for correctly writing out the definition of the left-hand side of Equation 1,
  - 1 point for correctly writing out the definition of the right-hand side of Equation 1,
  - 2 points for the derivation in Equation 4

  for either the case $(A+B)C = AC+BC$ or $A(C+D) = AC+AD$.
- Points may be awarded in other cases as well, but you only get full marks if your answer is completely correct for at least one of the two cases: $(A+B)C = AC+BC$ or $A(C+D) = AC+AD$.

2 Supervised Learning Problems

Exercise 2. *Give an example (not from class or from the book) of each of the following learning problems:*

- *prediction,*
- *classification,*
- *regression.*

How would you represent the labels and feature vectors (if applicable) mathematically? For either the classification or regression case (your choice), give at least two possible representations of your feature vectors.

Your examples do not need to be very realistic, but they do need to be about objects in the real world. For example, the answer "let $\mathbb{R}^6$ be the set of possible feature vectors" is not sufficient. You need to say what the feature vectors represent.

Answer Exercise 2: See the slides of the first three lectures for examples.

- Slides 23 and 24 of `mlslides1.pdf` provide examples of prediction that relate to phenomena in the real world.
- Slides 29 and 30 of `mlslides2.pdf` provide classification examples. Slide 31 provides a regression example.
- Slides 15 and 16 of `mlslides3.pdf` provide a classification example of representing the available data in two different ways using feature vectors.

Grading of Exercise 2: For examples of prediction, classification and regression in Exercise 2 you can get at most 1 point each. You get another 1 point for providing an alternative feature vector representation of the data in the classification or regression case. Thus you can get at most 4 points for this exercise.

3 Grading Policy

- Grades are between 1 and 10.
- You always start with 1 point.
- Partial points may be awarded for partially correct exercises.
- The last question of Exercise 2, about two possible representations of feature vectors, was perhaps unclear. It is therefore made a bonus question. If your grade according to the scoring above is x, then your final grade will be the minimum of $\frac{10}{9}x$ and 10.

4 Comment on Sources

In general webresources do not meet scientific standards. Nevertheless, I've consulted the ones below during the preparation of these exercises.

References

[1] J. Kolter. Linear algebra review and reference. Available online: `http://www.stanford.edu/class/cs229/section/cs229-linalg.pdf`, October 2006.