Solutions to selected problems from *Linear Algebra and Optimization for Machine Learning* (Ch. 1)

Luke Boyer

December 3, 2021

1.2 Scalars, Vectors, and Matrices

Exercise (Euclidean Distance as Difference)

Problem. *Show that* $||\bar{x} - \bar{y}||^2$ *is the squared euclidean distance between* $\bar{x}$ *and* $\bar{y}$*.*

Solution.

$$||\bar{x} - \bar{y}||^2 = (\bar{x} - \bar{y}) \cdot (\bar{x} - \bar{y})$$

$$z = [(x_1 - y_1), (x_2 - y_2) ... (x_d - y_d)]$$

$$z \cdot z = \sum_{i=1}^{d} (x_i - y_i)^2$$

This is the definition of euclidean distance squared.

□

Problem. *Solution.* □

Problem 1.2.1 (Triangle Inequality)

Problem. *Show that* $||\bar{x} - \bar{y}|| \leq ||\bar{x}|| + ||\bar{y}||$ *using cauchy's inequality.*

Solution. Recall that cauchy's inequality says that: $|\bar{x} \cdot \bar{y}| \leq ||\bar{x}|| ||\bar{y}||$.

Notice that since both sides of the equation are non-negative, it holds if and only if after squaring both sides.

$$(||\bar{x} + (-\bar{y})||)^2 = (\bar{x} + (-\bar{y})) \cdot (\bar{x} + (-\bar{y}))$$

$$= \bar{x} \cdot \bar{x} + 2\bar{x} \cdot (-\bar{y}) + \bar{y} \cdot \bar{y}$$

$$= ||\bar{x}||^2 - 2\bar{x} \cdot (-\bar{y}) + ||\bar{y}||^2 \leq ||\bar{x}||^2 + 2||\bar{x}|| ||\bar{y}|| + ||\bar{y}||^2 = (||x|| + ||y||)^2$$

Notice big step is applying Cauchy in the middle term of the last line and noticing it can be factored.

□

Problem 1.2.2 (Outer Product Properties)

Problem. *Show that the outer product of a* $n \times 1$ vector and a $1 \times d$ vector is a $n \times d$ *vector such that every row is a multiple of every other row and every column is a multiple of every other column.*

Solution. The dimensions of the result follow from the definition of matrix multiplication.

The i^{th} row of the resulting matrix looks like $[x_i y_1, x_i y_2 ... x_n y_d]$, thus every row is a multiple of $\bar{y}$ and they are all multiples of each other. Symmetrical reasoning holds for the columns. □

Problem 1.2.3

Problem. *Show that the product of matrices ABC can be expressed as a weighted sum of outer products of matrices vectors taken from A, C where the weights are taken from B.*

Solution. Let $A \in \mathbb{R}^{n\times d}$, $B \in \mathbb{R}^{d\times s}$, $C \in \mathbb{R}^{s\times k}$. Additionally let $\otimes$ be the outer product operator. For any matrix Z let $Z_{i,*}$ denote the i^{th} row of Z and similar for the j^{th} column but other ways around.

By previous lemma stated in the book and associativity we have that:

$$ABC = A(BC) = A\sum_{r=1}^{s} B_{*,r} \otimes C_{r,*}$$

Next we can distribute A over the summation and apply the same lemma to the product of A and $B_{*,r} \otimes C_{r,*}$ which is a $d \times k$ matrix.

$$= \sum_{r=1}^{s} A(B_{*,r} \otimes C_{r,*}) = \sum_{r=1}^{s}\sum_{t=1}^{d} A_{*,t}(B_{*,r} \otimes C_{r,*})_{t,*}$$

For a given step of the summation, notice that the entry (i,j) of the matrix resultant from the product inside the summation is given by $A_{i,t}B_{t,r}C_{r,j}$. The term from B is the same for every entry, thus it is equivalent to multiplication by a scalar, which commutes. After pulling the scalar notice that $A_{i,t}C_{r,j}$ is equivalent to the respective entry of $A_{*,t} \otimes C_{r,*}$. Thus, we may rewrite as:

$$\sum_{r=1}^{s}\sum_{t=1}^{d} B_{t,r}(A_{*,t} \otimes C_{r,*})$$

□

Problem 1.2.5

Problem. *Let $D \in \mathbb{R}^{n\times d}$ be a matrix whose columns sum to zero. Let $A \in \mathbb{R}^{d\times d}$ be any matrix. Show that the product of DA also has columns summing to zero.*

Solution. We can examine the i^{th} column of DA with the same column/row selection notation as above.

$$\begin{bmatrix} D_{1,*} \cdot A_{*,i} \\ D_{2,*} \cdot A_{*,i} \\ . \\ . \\ D_{n,*} \cdot A_{*,i} \end{bmatrix}$$

Thus the sum of the i^{th} column can be written as:

$$\sum_{j=1}^{n} D_{j,*} \cdot A_{*,i}$$

The inner product distributes and thus we can rewrite as:

$$(\sum_{j=1}^{n} D_{j,*}) \cdot A_{*,i}$$

Notice that by definition the i^{th} component of $(\sum_{j=1}^{n} D_{j,*})$ is the sum of the i^{th} column of D and thus $(\sum_{j=1}^{n} D_{j,*})$ is just a row vector of zeros implying $(\sum_{j=1}^{n} D_{j,*}) \cdot A_{*,i}$ is just the scalar zero. □

Problem 1.2.8 (Inverse of Triangular Matrix is Triangular)

Let R be a $d \times d$ upper triangular matrix and $\bar{e_k}$ be a d dimensional vector with 1 in the k^{th} entry and 0 elsewhere. Discuss why solving $\bar{x}$ is simple in the equation $R\bar{x} = \bar{e_k}$.

For any equation $R\bar{x} = b$, and upper triangular matrix R it is the case that $R_{d,d}\bar{x}_d = \bar{b_d}$ which we can solve. With this information we can solve for x_{d-1} similarly in an equation with one unknown and repeat the process for all components of $\bar{x}$.

Discuss why $R\bar{x} = \bar{e_k}$ must satisfy $x_i = 0$ for $i > k$.

Since R is triangular, the entries on the diagonal must be non-zero. If $k \neq d$, then x_d must be zero. We can apply the same reasoning as above to conclude that all the entries of $\bar{x}$ from d to $k+1$ must be zero.

Discuss why the solution to $\bar{x}$ is equal to the k^{th} column of the inverse of r.

It is easy to see for any matrix B, $B\bar{e_k}$ is the k^{th} column of B. So $R\bar{x} = \bar{e_k} \rightarrow \bar{x} = R^{-1}\bar{e_k}$.

Discuss why the inverse of R is upper triangular.

We can expand our reasoning from above and replace x with a $d \times d$ matrix and $\bar{e_k}$ with the identity matrix. Notice that finding the k^{th} column of the new matrix equates to solving the original equation with $\bar{e_k}$. It follows from the previous sections that the all components of the k^{th} column that come after the k^{th} component are zero, yielding the desired upper triangular structure.

Problem 1.2.11 (Inverting Triangular Matrices)

coming soon

Problem 1.2.12

Problem. *Suppose that I and K are two $d \times d$ matrices. Show that:*

$$(I+P)^{-1} = I - (I+P)^{-1}P$$

Solution.

$$I - (I+P)^{-1}P = (I+P)^{-1}[(I+P) - P]$$

$$= (I+P)^{-1}[I] = (I+P)^{-1}$$

□

Problem 1.2.13 (Push Through Identity)

Problem. *Show:* $U^T(I_n + VU^T)^{-1} = (I_d + U^T)^{-1}U^T$

Solution.

$$U^T(I_n + VU^T)^{-1} = (I_d + U^TV)^{-1}U^T \rightarrow$$

$$(I_d + U^TV)U^T = U^T(I_n + VU^T) \rightarrow$$

$$U^T + U^TVU^T = U^T + U^TVU^T$$

□

Problem. *Show:* $D^T(\lambda I_n + DD^T)^{-1} = (\lambda I_d + D^TD)^{-1}D^T$

Solution.

$$D^T(\lambda I_n + DD^T)^{-1} = (\lambda I_d + D^TD)^{-1}D^T \rightarrow$$

$$(\lambda I_d + D^TD)D^T = D^T(\lambda I_n + DD^T) \rightarrow$$

$$\lambda D^T + D^TDD^T = \lambda D^T + D^TDD^T$$

□

Problem 1.2.14

Problem. *Show that the Frobenius norm of the outer product of two vectors is equal to the product of their Euclidean norms.*

Solution. Let s, t be $n-$dimensional vectors. Recall that entry (i, j) of $s \otimes t$ is $s_i t_j$.

$$\|s \otimes t\|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (s_i t_j)^2}$$

$$= \sqrt{\sum_{i=1}^{n} s_i^2 (\sum_{j=1}^{n} t_j^2)} = ||t|| \sqrt{\sum_{i=1}^{n} s_i^2} = \|t\| \|s\|$$

□

Problem 1.2.15 (Small Matrices Have Large Inverses)

Problem. *Show that the F-norm of the inverse of an $n \times n$ matrix with F-norm of ϵ is at least* $\frac{\sqrt{n}}{\epsilon}$

Solution. Let A be an $n \times n$ matrix that is non-singular. Notice that $\|I_n\|_F = \sqrt{n}$. One way to see this is through the identity $\|D\|_F^2 = trace(DD^T)$ for any matrix D and that the identity is symmetric. Now consider:

$$AA^{-1} = I_n \implies \|AA^-1\|_F = \|I_n\|_F = \sqrt{n}$$

By sub-multiplicity of the F norm we have that

$$\sqrt{n} = \|AA^-1\|_F \leq \|A\|_F \|A^{-1}\|_F = \epsilon \|A^{-1}\|_F$$

$$\implies \frac{\sqrt{n}}{\epsilon} \leq \|A^{-1}\|_F$$

□

Problem 1.5.1

Problem. *Show using Taylor expansions that* $e^{i\theta} = \cos\theta + i\sin\theta$

Solution. Consider the Taylor expansion of $e^{i\theta}$ about $\theta = 0$ which is given by:

$$\sum_{j=0}^{\inf} \frac{i^j \theta^j}{j!}$$

Notice that the even terms of this series are equal to the cosine expansion and the odd terms are equal to the sin expansion. □ □

Exercise 2.5 (Sine law)

This is work in progress

Problem. *Express the sine of the interior angle between two vectors a, b only in terms of dot products.*

Solution. One way to see the sine is the ratio between the length of b and the length of the line drawn from b to some point on a that is orthogonal to a. We can compute the vector for this line by taking the difference of b and b projection onto a. Note that the projection of b onto a is given by $(\frac{a \cdot b}{a \cdot a})a$. Thus,

$$\sin\theta = \frac{\|b - (\frac{a \cdot b}{a \cdot a})a\|}{\|b\|}$$

□

Exercise 26 (Trigonometry with vector algebra)

Coming soon

Exercise 27 (Coordinate geometry with matrix algebra)

Coming soon

Exercise 29 (Solid geometry with vector algebra)

Coming soon

Exercise 31 (Matrix centering)

Coming soon

Exercise 49 (Inverses behave like matrix polynomials)

Coming soon

Exercise 32 (Energy preservation in orthogonal transformations)

Problem. *If A is an $n \times d$ matrix and P and an orthonormal $d \times d$ matrix, then $\|A\|_F = \|AP\|_F$.*

Solution. Multiplication by orthonormal matrices preserve the magnitude of an input vector. An arbitrary row of the result AP is equal to the corresponding row of A pre-multiplied with the orthonormal P. Thus, the rows of AP have the same magnitude as their corresponding row in A, i.e. $\|A\|_F^2 = \|AP\|_F^2$ □

Exercise 33 (Tight sub-multiplicative case)

Problem. *Suppose that u, v are column vectors with dimensions n, d respectively. Show that the F-norm of their outer product is equal to $\|u\|\|v\|$.*

Solution. Notice that entry (i, j) of $u \otimes v$ is $u_i v_j$.

$$\|u \otimes v\|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{d} u_i^2 v_j^2}$$

$$= \sqrt{\sum_{i=1}^{n} u_i^2 \sum_{j=1}^{d} v_j^2} = \|v\| \sqrt{\sum_{i=1}^{n} u_i^2} = \|v\|\|u\|$$

□

Exercise 24 (Frobenius orthogonality and Pythagorean theorem)

Problem. *Two $n \times d$ A, B are said to be Frobenius orthogonal if the sum of the entry-wise product of their corresponding elements is zero. Show that $\|A+B\|_F^2 = \|A\|_F^2 + \|B\|_F^2$*

Solution.

$$\|A+B\|_F^2 = tr((A+B)(A+B)^T) = tr((A+B)(A^T+B^T))$$

$$= tr(AA^T + A^TB + AB^T + BB^T) = tr(AA^T) + tr(A^TB) + tr(AB^T) + tr(BB^T)$$

$$= tr(AA^T) + tr(BB^T) = \|A\|_F^2 + \|B\|_F^2$$

□