Deep Dive Problems

1. *Revisiting the least of your problems*(**#vectors**, **#linearsystems**) In this problem, we will revisit the least-squares approximation to a system of linear equations introduced in Assignment 2.

   (a) Consider a system $A\mathbf{x} = \mathbf{b}$. In what fundamental subspace of A must the right-hand side vector $\mathbf{b}$ be, in order for the system to have at least one solution?

   Answer:

   For $A\mathbf{x} = \mathbf{b}$ to have at least one solution, $\mathbf{b}$ has to be a linear combination of columns of A.

   Let's assume that: $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}, x = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$.

   If we read the equation again we see that $\mathbf{b}$ is the product of A and $\mathbf{x}$. Thus, let's compute the multiplication:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$x \begin{bmatrix} a \\ d \\ g \end{bmatrix} + y \begin{bmatrix} b \\ e \\ h \end{bmatrix} + z \begin{bmatrix} c \\ f \\ i \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

   Now we see that we have expressed $\mathbf{b}$ as a linear combinations of columns in A. This means that $\mathbf{b}$ is in the **column space** of A, because each linear combination of vectors in a space spans that space.

   (b) If we are given that the system $A\mathbf{x} = \mathbf{b}$ has no solution, then $\mathbf{b}$ must have a component in which fundamental subspace of A?

   Answer:

   We see that we have inconsistencies in the system when $A\mathbf{x} = \vec{0}$ and $\mathbf{b} \neq \vec{0}$.

   - For $A\mathbf{x} = \vec{0}$, either A or $\mathbf{x}$ should be $\vec{0}$.
   - For $\mathbf{b} \neq \vec{0}$, at least one of the components b_1, b_2, b_3 has to be non-zero.
     If we take $\mathbf{x} = \vec{0}$ and $b_1 \neq \vec{0}$, we are sure to have an inconsistency in the system, hence making it not solvable.
     Let's observe it:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} b_1 \\ 0 \\ 0 \end{bmatrix}$$

After we multiply, we get:

$$0 \begin{bmatrix} a \\ d \\ g \end{bmatrix} + 0 \begin{bmatrix} b \\ e \\ h \end{bmatrix} + 0 \begin{bmatrix} c \\ f \\ i \end{bmatrix} = \begin{bmatrix} b_1 \\ 0 \\ 0 \end{bmatrix}$$

If we continue with the computation, we end up with the following equations:

$$0a + 0b + 0c = b_1$$
$$0d + 0e + 0f = 0$$
$$0g + 0h + 0i = 0$$

In the first equation, we observe that the left hand-side has to be zero, because it is the sum of products of 0, but in the right-hand side we have a non-zero component. This means that the system is unsolvable. If $b_1 = 0$, we would have infinite solutions to the system.
So, we've created an inconsistent system by changing one component of $\mathbf{b}$ into a non-zero element, while making the linear combination of A equal to 0. Since for the system to have a solution we decided that $\mathbf{b}$ should be in the column space, for the system to be inconsistent we find that at least one component has to be part of the orthogonal space to column space, which is the left-nullspace.

(c) The previous two parts imply that the vector $\mathbf{b}$ may be decomposed as $\mathbf{b} = \mathbf{v_a} + \mathbf{v_b}$, where $\mathbf{v_a}$ is the component of $\mathbf{b}$ in the fundamental space from part (a), and $\mathbf{v_b}$ is the component of $\mathbf{b}$ in the fundamental space from part (b). What is the relationship between $\mathbf{v_a}$ and $\mathbf{v_b}$?

Answer:

$\mathbf{v_a}$ refers to the column space vector, while $\mathbf{v_b}$ lies in the left-nullspace. We saw that those two vectors should not be contained together into $\mathbf{b}$, otherwise the system has no solutions. Moreover, the relationship between these two subspaces (and each component within them) must be orthogonal, for a solvable solution. As we know, similar orthogonality pattern is observed between rowspace and nullspace. The dot product of the vectors of such spaces must be equal to 0, thus $\mathbf{v_a} \cdot \mathbf{v_b} = 0$.

(d) Our goal now is to find a vector $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}}$ minimizes the distance from $\mathbf{b}$. How does $A\hat{\mathbf{x}}$ relate to the vectors $\mathbf{v_a}$ and $\mathbf{v_b}$ above?

i. Recall that, in assignment 2, you derived that

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \vec{\mathbf{0}}.$$

Interpret this statement in terms of the fundamental subspaces of A.

Answer:

What this question implies is that $\mathbf{b}$ is already not in the column space, but we can project it onto column space of A closely to get an approximation of $\mathbf{x}$, that we denote as $\hat{\mathbf{x}}$. We want to minimize the difference between $\mathbf{b}$ and its projection into $col(A)$. This difference can also be denoted as $\mathbf{b}-\mathrm{A}\hat{\mathbf{x}}$. If $\mathbf{b}$ was in the column space of A, then this difference would be 0.

This equation

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \vec{\mathbf{0}}.$$

resembles with the definition of left-nullspace which is

$$A^T y = \vec{\mathbf{0}}.$$

Thus, we can see that $(\mathbf{b} - A\hat{\mathbf{x}})$ represents the y. Now, we can say that y is going to be the difference between $\mathbf{b}$ and its projection onto column space A, because $A\hat{\mathbf{x}}$ is represented in the column space, while b has a component in the left-nullspace.

(e) Prove the following statement: If the columns of A are linearly independent, then A^TA is invertible. (Hint: Prove that $N(A) = N(A^TA)$. What does that mean about the rank of the matrices?)

Answer:

Some things to keep in mind:

i. If the columns are linearly independent, then the only solution to the homogeneous system is the trivial solution($x = 0$).

ii. An invertible matrix when multiplied with its inverse should produce the identity matrix. The matrix has to be **nxn**.

iii. $A\mathbf{x} = 0$ represents the nullspace of A.

iv. $Rank(\mathrm{A}) = dim(row(A)) = dim(col(A))$

v. If all columns are linearly independent, the nullspace and left-nullspace are 0, because the matrix is **nxn**, and $dim(row(A))+dim(N(A)) = n$ and $dim(col(A))+dim(N(A^T)) = n$. Considering point 4, we see that $Rank(A) = n$, thus $dim(N(A)) = 0$ and $dim(N(A^T)) = 0$.

Linear independence can be represented as a linear combination that equals 0, if and only if $\vec{x} = \vec{0}$. This means that the $N(A) = \vec{0}$.

Is $N(A) = N(A^TA)$?

As we said, x lies in the nullspace of A, and if we assume that x lies in $N(A^TA)$, then the following equations must be met $A^TAx = 0$. If we left-multiply by the tranpose of x, we get $x^TA^TAx = 0$, than can be written as $(Ax)^TAx = 0$. If we rewrite Ax as a, we end up with $a^Ta = 0$.

$a^T a = 0$ means that the multiplication of a matrix with its own transpose will be 0. This is only possible if all the entries of a are 0.

Thus, $a^T a = a \cdot a = Ax \cdot Ax = 0$

Since Ax is multiplying itself and the product is 0, then $Ax = 0$, which represents the nullspace of A. From the linear independence rule above, we asserted that the only solution to the homogeneous system is the trivial solution, which means $x = 0$, thus $N(A^T A)$ only contains one vector which is the zero vector. **This proves linear independence**.

Invertibility requires the matrix to be square.

$A^T A$ **is a square matrix** because each matrix that multiplies its transpose turns into a square matrix (since the number of columns of A, must be equal to the number of rows in A^T, which then multiply each other.

Hence, we have proven that $A^T A$ is invertible.

(f) Finally, we can manipulate the above formula to solve for $\hat{x}$. Show that

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}.$$

Answer:

We can expand this equation further:

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \vec{\mathbf{0}}.$$

$$A^T\mathbf{b} - A^T A\hat{\mathbf{x}} = \vec{\mathbf{0}}.$$

And we can balance both sides of the equation:

$$A^T\mathbf{b} = A^T A\hat{\mathbf{x}}$$

If we want to get $\hat{\mathbf{x}}$ from here, then we need to left-multiply with $(A^T A)^{-1}$, knowing that $A^T A$ is invertible.

$$(A^T A)^{-1} A^T\mathbf{b} = I\hat{\mathbf{x}}$$

$$(A^T A)^{-1} A^T\mathbf{b} = \hat{\mathbf{x}}$$

(g) Regression Example. While visiting their family last weekend, your friend's nine-year-old sister asked you for help with her science project. She spent the afternoon at a small, local pond counting the number newt eggs that hatched. Here is a table of her data:

The data is missing between 2 pm and 3 pm when she took a break to play on a tire swing by the pond. She's creating a plot of her data. She really likes math, so you figured you could find the regression line and show her how to use the line to estimate how many eggs hatched during her break.

Time of measurement	Eggs Hatched
Noon	0
1 pm	6
3 pm	8
4 pm	18

i. Use the data to create a system of equations in two unknowns - C and M. Write out this system explicitly, then describe it as the matrix equation $Ax = b$, where $x = \begin{bmatrix} C \\ M \end{bmatrix}$. What is A?
Answer:
We need to see what equations we need to fulfill in order to estimate future predictions. In that case we try to represent the eggs hatched as a function of time, but since this is a prediction, there is an error rate that needs to be taken into account.
The equations takes the shape of

$$Eggshatched = predictionerror + (afactor) * time.$$

The prediction error is constant, thus we will name it as C, which comes from the distance of data points from the best-fit line. If the data fits perfectly in the line, then we have $C = 0$ (no distance between data and the line), but usually not each function has the form of perfect linear function. That is because the *eggshatched* are a function of other factors but *time*, that we do not take into account, or simply because the nature of the relationship is not linear. *Time* is just one predictor of many, which makes this a simple regression model. However, from the calculation of Pearson's R (code in R-Appendix), we see that *time* is quite correlated (0.9917) with *eggshatched*, which means that we can explain the variability in *eggshatched* by 99.17% only taking *time* as a factor, thus the prediction error will be small. [1] The other way we can see prediction error is that, some phenomena is complex, rather than complicated, which means that the outcome is more than the sum of its parts. For instance, even if we knew all the factors and their quantitative contribution to *eggshatched*, we may not be able to predict it perfectly because the phenomenon of eggs being hatched is complex, which is different from things that are only complicated (i.e.the outcome of a computer function).[2] By knowing the error,

[1] #correlation: Used correlation to explain the concept of x as a predictor and y as a predicted value; used a computational tool to plug in the value of Pearson's R.

[2] #emergentproperties: Explained the difference between a complicated and complex system in terms of

we improve our estimation of prediction for the future.The factor of time is also a constant which will always multiply time, and we can denote it as M.
We can write the equation from data as follows:

$$C + 0M = 0$$

$$C + 1M = 6$$

$$C + 3M = 8$$

$$C + 4M = 18$$

We take time as 0-the beginning, 1-first time step, we skip 2-second time step as we have no data for it, 3- third time step, 4- fourth time step. In matrix form, this can be represented as:

$$A\vec{x} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} C \\ M \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 8 \\ 18 \end{bmatrix} = b$$

ii. Demonstrate that the system is inconsistent and that the columns of A are independent.
Answer:
To prove linear independence and inconsistencies, we turn the given matrix of coefficients into Row Reduced Echelon Form, by augmenting it with the solutions.

$$\begin{bmatrix} 1 & 0 & | & 0 \\ 1 & 1 & | & 6 \\ 1 & 3 & | & 8 \\ 1 & 4 & | & 18 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & | & 0 \\ 0 & 1 & | & 0 \\ 0 & 0 & | & 1 \\ 0 & 0 & | & 0 \end{bmatrix}$$

Both column 1 and 2 have pivots of 1 and all other elements are 0, which proves that these two columns are linearly independent. However in row 3, we find the equation $0C + 0M = 1$, which in our scenario means that we will get 1 egg hatched, without time even starting. This is consistent with the data that we have, as the first row in the table shows exactly the same thing, but it is inconsistent mathematically, because there is no other number than 0, as a sum of two zero products. This is why we say that, the system is inconsistent.[3]

regression prediction model and egg hatching example. Made sure to call out that egg hatching is complex and its outcome is more than the sum of its parts.

[3] All calculations were made in Sage.

iii. Use the techniques above to solve for $\hat{x}$.
<u>Answer:</u>

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}.$$

We start by replacing the known values.

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \xrightarrow{transpose} A^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix}$$

$$(A^T A)^{-1} = \begin{bmatrix} \frac{13}{20} & \frac{-1}{5} \\ \frac{-1}{5} & \frac{1}{10} \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 6 \\ 8 \\ 18 \end{bmatrix} = \begin{bmatrix} 32 \\ 102 \end{bmatrix}$$

$$(A^T A)^{-1} A^T b = \begin{bmatrix} \frac{2}{5} \\ \frac{19}{5} \end{bmatrix} = \begin{bmatrix} 0.4 \\ 3.8 \end{bmatrix} = \hat{x}$$

The results are consistent with the regression model outputted from R (R-code Appendix).

iv. Make a diagram that includes each of the data points and the regression line.
<u>Answer:</u>

v. Write a short explanation of the result aimed at your nine-year-old audience. Make sure to include an estimate for the number of eggs hatched while she was away.
<u>Answer:</u>
We believe that the eggs being hatched is not a spontaneous process; it does not just happen, but it follows a pattern/a trend. Let's assume it is a straight-line trend (which in mathematics, we call it linear). So we try to find the math formula for the trend based on what information we have.
As we know, you observed the eggs hatching process for about 5 hours, while skipping one time step. What we know from your data is that the number of eggs hatched at 2pm cannot be less than 6 or bigger than 8, because you had 6 at 1pm and 8 at 3pm. So, it can be 6, 7 or 8. When we see the relationship between the time that you gave me and the numbers of eggs, we see that if time is multiplied by 3.8 and we add 0.4 to that product, we will get the approximate number of eggs being hatched. A big emphasis

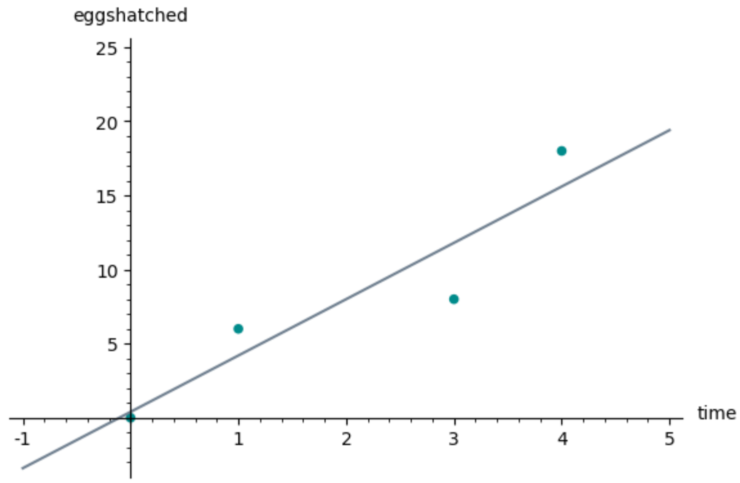

Figure 1: The regression line and data points.

should be put in approximate, because that does not mean you will get the exact same number as the number of eggs hatched in the real life, for example $3.8 * 1 + 0.4 = 4.2$, while at 1pm there were 6 eggs. So there must be something except time that impacts the egg hatching, that we aren't aware of, but using time, we can make somewhat accurate predictions. So, let's plug-in the same formula at 2pm, when you went to the pond.

$$3.8 * 2 + 0.4 = 8$$

Thus, there is a high chance that there were 8 eggs that hatched while you were away. This is plausible, because 8 is not smaller than 6 and bigger than 8. [4]

(h) Fitting a curve. Your protégé was not particularly impressed. She thought the line was "too far" from her points to make an accurate estimate. So you decide to fit a parabola and a cubic so she can compare the results.

i. For the closets parabola $b = C + Dt + Et^2$ to the same four points:

A. Write the system of equations $Ax = b$ in unknowns $x = (C, D, E)$. Show the system is inconsistent and the columns of A are independent.
Answer:
Similarly, we write the equations with the given formula:

$$b = C + Dt + Et^2$$

[4] #regression: Thoroughly explained the least of squares, while applying regression concepts. Ran a regression model in R to check the results which we computed by the derived formula. Interpreted regression in the context of egg hatching.

$$0 = C + 0D + 0E$$
$$6 = C + D + E$$
$$8 = C + 3D + 9E$$
$$18 = C + 4D + 16E$$

In the matrix form, this looks like:

$$A\hat{x} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 8 \\ 18 \end{bmatrix}$$

In RREF, matrix A looks like:

$$A = \begin{bmatrix} 1 & 0 & 0 & | & 0 \\ 1 & 1 & 1 & | & 6 \\ 1 & 3 & 9 & | & 8 \\ 1 & 4 & 16 & | & 18 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & 0 & | & 0 \\ 0 & 1 & 0 & | & 0 \\ 0 & 0 & 1 & | & 0 \\ 0 & 0 & 0 & | & 1 \end{bmatrix}$$

We see that all columns in the matrix have pivots of 1 and 0s in the other components, thus the columns are linearly independent. The last row represents $0C + 0t + 0t^2 = 1$, which is impossible because no sum of 0 product can be more than 0 itself. With this absurd equation, we can conclude that the system is inconsistent.

B. Find $\hat{x}$. Describe the geometric relationship between A, $\hat{x}$, and b.
<u>Answer:</u>
After we replace everything in the given formula $(A^TA)^{-1}A^T\mathbf{b} = \hat{\mathbf{x}}$, we get

$$\hat{\mathbf{x}} = \begin{bmatrix} \frac{7}{5} \\ \frac{17}{5} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 1.4 \\ 3.4 \\ 0.6666 \end{bmatrix}$$

This means that the parabola equation will be:[5]

$$1.4 + 3.4t + 0.6666t^2 = b$$

Given this formula we see that the relationship of b with $A\hat{x}$ is parabolic of the grade 2, which means that b is explained by the square of $\hat{x}$.

C. Add the parabola determined by $\hat{x}$ to your plot. Analyze the result. Do you expect this to give a better prediction?
<u>Answer:</u>

[5] The number rounded to the 4th decimal.

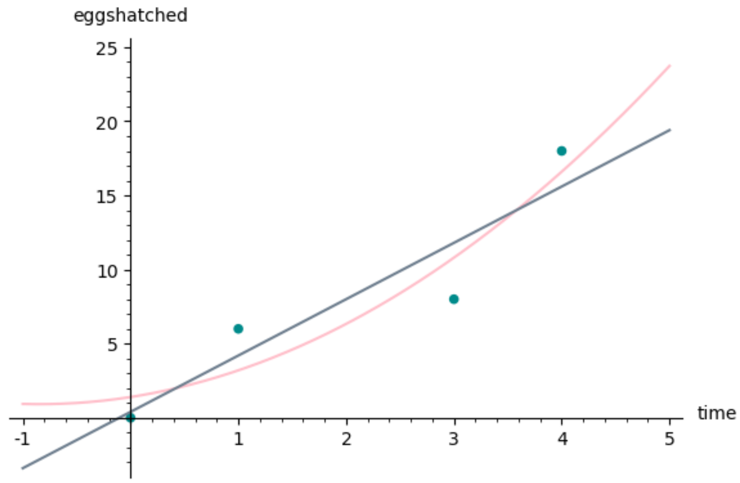

Figure 2: The parabola and regression line.

We see from the graph that the parabola has less distance with the data points, except for point 1 and 2 that regression model is closer for. This seems to be a closer approximation to the real trend, but that does not mean that this model is better/worse than the regression model. A detail that shows the ineffectivity of results of this model is that for negative values of time it shows positive number of eggs hatched. This is nonsensical for the real data, although we are not particularly interested for the negative values of time as that does not exist in real conditions.

ii. Write the system for the best cubic $b = C + Dt + Et^2 + Ft^3$ and try to repeat the process. What happens this time? Add the best fit cubic to your plot. Do you expect this to be a more effective model? Why or why not?

Answer:

We rewrite the equation:

$$b = C + Dt + Et^2 + Ft^3$$

$$0 = C + 0D + 0E + 0F$$

$$6 = C + D + E + F$$

$$8 = C + 3D + 9E + 27F$$

$$18 = C + 4D + 16E + 64F$$

In matrix form this looks like:

$$A\hat{x} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \\ F \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 8 \\ 18 \end{bmatrix}$$

In RREF, matrix A looks like:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & | & 0 \\ 1 & 1 & 1 & 1 & | & 6 \\ 1 & 3 & 9 & 27 & | & 8 \\ 1 & 4 & 16 & 64 & | & 18 \end{bmatrix} \xrightarrow{RREF} \begin{bmatrix} 1 & 0 & 0 & 0 & | & 0 \\ 0 & 1 & 0 & 0 & | & \frac{67}{6} \\ 0 & 0 & 1 & 0 & | & \frac{-19}{3} \\ 0 & 0 & 0 & 1 & | & \frac{7}{6} \end{bmatrix}$$

Since all columns have pivots, and all other entries are 0, we say the columns are linearly independent. We see no inconsistencies in the system, but instead we find the values of $\hat{x}$.

$$C = 0$$

$$D = 11.1666$$

$$E = -6.3333$$

$$F = 1.1666$$

The line equation is:

$$11.1666t - 6.3333t^2 + 1.1666t^3 = b$$

Since C is zero, which we earlier called the prediction error, then we have no error, thus we expect the line to meet all the data points.
Let's examine the plot:
This model effectively meets all the data points, thus it has no residual error (prediction error). We might have found the pattern of egg hatching and since it fits perfectly in the given data, it may fit further in time as well. We cannot be sure that this is the pattern, thus, the predictions for the future may still be inaccurate. We have only proved that the relationship between time and numbers of egg hatched for the first time steps is cubic.

iii. Write a short explanation of the results of curve fitting aimed at your nine-year-old audience.
Answer:
As we said earlier, we assume there is a trend. If we do not think that the trend is a straight-line (linear), the trend may be parabolic, which shows this curve in the Figure 2, or the trend may be cubic like Figure 3. For the missing observation at 2pm, the straight-line model says that there were

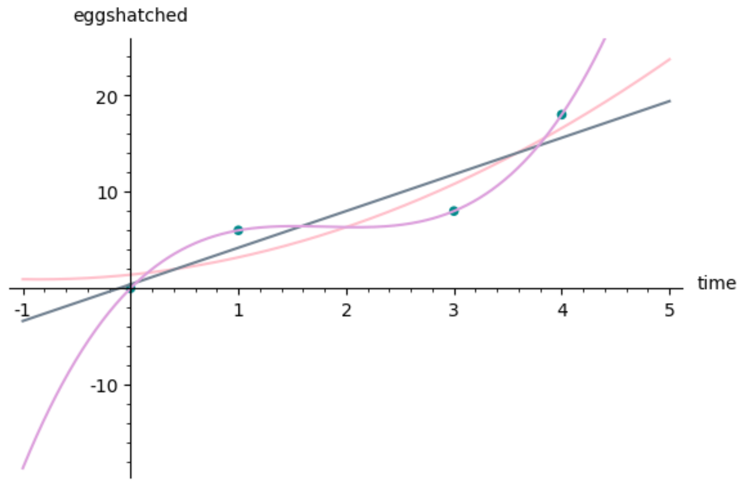

Figure 3: The cubic line, parabola and regression line.

8 eggs hatched, the parabolic model shows 10.8664 hatched, and the cubic 6.33. Knowing the constraint there were either 6,7 or 8 eggs hatched, we can eliminate the predictions that the parabolic model gives, because 10.8864 is bigger than 8.
But which model do we trust regression or cubic? For sure, it seems odd to have 6.33 eggs as we count eggs as a whole number, not decimal. But that is not enough reason not to trust the cubic model; just because the regression model is outputting a whole number, it does not mean the prediction is closer to the actual number of eggs hatched. We see that the cubic model maps perfectly into the line of 5 first time steps, thus we have more reasons to trust the cubic model. Does that mean that the model will always predict accurately? No, because we are not sure that the number of eggs hatched can be predicted by the passing time only. So, 6.33 eggs, is how many eggs exactly? It is either 6 or 7, but most likely 6.
So, with a degree of small uncertainty, we can say that there were 6 eggs hatched at 2pm. Seeing that the models cannot perfectly predict the real world phenomenon, you should stay actively by the pond and observe for 100% guaranteed accuracy and certainty.

Novel Problems

1. *Inside Information*(**#vectors**, **#theoreticaltools**) Dot product is an example of the more general notion of an *inner product*. An *inner product* on a vector space V

is an operation that takes as input a pair vectors $v, w \in V$ and outputs a number $\langle v, w\rangle \in R$ satisfying the following conditions:

- *Positivity*: $\langle v, v\rangle \geq 0$ for all $v \in V$
- *Definiteness*: $\langle v, v\rangle = 0$ if and only if $v = \vec{0}$ for all $v \in V$.
- *Additivity*: $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$
- *Homogeneity*: $\langle av, w\rangle = a\langle v, w\rangle$ for all $a \in \mathbb{R}$, $v, w \in V$
- *Symmetry*: $\langle v, w\rangle = \langle w, v\rangle$ for all $v, w \in V$

(a) Show that dot product on $\mathbb{R}^n$ fulfills the conditions above.

Answer

Positivity: $\langle v, v\rangle \geq 0$ for all $v \in V$

If $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$, we can compute the dot product of $\vec{v} \cdot \vec{v}$ as

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v_1^2 + v_2^2 + \cdots + v_n^2$$

For any values of the components in $\vec{v}$, we will get positive values, except 0, as the square of any real numbers is positive and the square of 0 is 0.

Definiteness: $\langle v, v\rangle = 0$ if and only if $v = \vec{0}$ for all $v \in V$.

Let's take a non-zero vector and see if we can reach to zero product:

$$\vec{v} \cdot \vec{v} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ \vdots \\ v_n \end{bmatrix} = 0 + 0 + \cdots\ + v_n^2 \neq 0$$

We see that even if only one component is non-zero, we reach to the conclusion that we cannot produce the $\vec{0}$ vector, because one value will be non-zero (v_n^2), as there is no such scalar that when it multiplies itself it produces 0.

Additivity: $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$

$$\langle u + v, w\rangle = [\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}] \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} v_1 + u_1 \\ v_2 + u_2 \\ \vdots \\ v_n + u_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} =$$

$$= v_1w_1 + u_1w_1 + v_2w_2 + u_2w_2 + \cdots + v_nw_n + u_nw_n$$

$$\langle u, w\rangle + \langle v, w\rangle = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} v_1w_1 \\ v_2w_2 \\ \vdots \\ v_nw_n \end{bmatrix} + \begin{bmatrix} u_1w_1 \\ u_2w_2 \\ \vdots \\ u_nw_n \end{bmatrix} =$$

$$= v_1w_1 + u_1w_1 + v_2w_2 + u_2w_2 + \cdots + v_nw_n + u_nw_n$$

We have proven that right-hand side is equal to the left-hand side.
Homogeneity: $\langle av, w\rangle = a\langle v, w\rangle$ for all $a \in \mathbb{R}$, $v, w \in V$

$$\langle av, w\rangle = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} a_1v_1 \\ a_2v_2 \\ \vdots \\ a_nv_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} =$$

$$= a_1v_1w_1 + a_2v_2w_2 + \cdots + a_nv_nw_n$$

$$a\langle v, w\rangle = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} v_1w_1 \\ v_2w_2 \\ \vdots \\ v_nw_n \end{bmatrix} =$$

$$= a_1v_1w_1 + a_2v_2w_2 + \cdots + a_nv_nw_n$$

We have proven that the right-hand side is equal to the left-hand side.
Symmetry: $\langle v, w\rangle = \langle w, v\rangle$ for all $v, w \in V$

$$\langle v, w\rangle = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = v_1w_1 + v_2w_2 + \cdots + v_nw_n$$

$$\langle w, v\rangle = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = w_1v_1 + w_2v_2 + \cdots + w_nv_n$$

As the product is now scalar product, we know that the symmetry property holds $ab = ba$, which is equivalent to $v_nw_n = w_nv_n$, in our scenario. Thus, we have proven the left-hand side to be equivalent to right-hand side.

(b) Show that $\langle A, B\rangle = tr(A^T B)$ is an inner product on $\mathcal{M}_{2\times 2}$, where the trace of a matrix is defined as the sum of the entries along the main diagonal.

Answer:

If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $B = \begin{bmatrix} \alpha & \beta \\ \gamma & \theta \end{bmatrix}$, then $\langle A, B\rangle = a\alpha + b\beta + v\gamma + d\theta$.

Now let's check the right-hand side of the equation:

$$tr(A^T B) = tr(\begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} \alpha & \beta \\ \gamma & \theta \end{bmatrix}) = tr(\begin{bmatrix} a\alpha + c\gamma & a\beta + c\theta \\ b\alpha + d\gamma & b\beta + d\theta \end{bmatrix}) = a\alpha + c\gamma + b\beta + d\theta$$

Hence, we proved the left-hand side and right-side of the equation.

(c) Using inner products we can define the *norm* (or length) of a vector as $\|v\| = \sqrt{\langle v, v\rangle}$. Derive a formula for the norm of a matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ using the inner product definition from part (b).

Answer:

From the previous question we proved that $\langle A, B\rangle = tr(A^T B)$.

Thus,

$$\langle M, M\rangle = tr(M^T M) =$$

$$= tr[\begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}] = tr[\begin{bmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{bmatrix}] = a^2 + b^2 + c^2 + d^2$$

Since the norm is defined by the square root of $\langle M, M\rangle$, then we take the square root of the last equation:

$$||M|| = \sqrt{a^2 + b^2 + c^2 + d^2}$$

(d) We can now extend the concept of orthonormal bases to other vector spaces. A basis $S = \{v_1, v_2, ..., v_n\}$ is *orthonormal* if $\langle v_i, v_i\rangle = 1$ and $\langle v_i, v_j\rangle = 0$ for $i \neq j$ for $1 \leq i, j, \leq n$. Show that the basis $\left\{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right\}$ of $\mathcal{M}_{2\times 2}$ is orthonormal under the inner product from part(b).

Answer:

Orthonormality requires:

- The inner product/norm should be 1.
- The inner product with other matrices must be 0.

We can check this by computing the inner product with the formula above which is practically

$$a_{ii}^2 + a_{ij}^2 + a_{ji}^2 + a_{jj}^2$$

We substitute and get:
First Matrix: $1 + 0 + 0 + 0 = 1$
Second Matrix: $0 + 1 + 0 + 0 = 1$
Third Matrix: $0 + 0 + 1 + 0 = 1$
Fourth Matrix: $0 + 0 + 0 + 1 = 1$
All the matrices have inner product of 1. Now we need to find that their dot product with one another is 0. That changes the above formula to:

$$a_{ii}b_{ii} + a_{ij}b_{ij} + a_{ji}b_{ji} + a_{jj}b_{jj}$$

Since dot product is commutative, we get the following combinations:
First and Second Matrix: $0 + 0 + 0 + 0 = 0$
First and Third Matrix: $0 + 0 + 0 + 0 = 0$
First and Fourth Matrix: $0 + 0 + 0 + 0 = 0$
Second and Third Matrix: $0 + 0 + 0 + 0 = 0$
Second and Fourth Matrix: $0 + 0 + 0 + 0 = 0$
Third and Fourth Matrix: $0 + 0 + 0 + 0 = 0$
This is how we've proved that those vectors form an orthonormal basis.

(e) Gram-Schmidt can also be extended. Apply the Gram-Schmidt procedure to convert the basis $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 3 & 0 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 6 \\ 0 & 0 \end{bmatrix} \right\}$ into an orthonormal basis.

Answer:

We will denote the original matrices as o_1, o_2, o_3, o_4 and orthonormalized matrices as n_1, n_2, n_3, n_4.

$$v_1 = o_1 = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

$$n_1 = \frac{v_1}{\sqrt{1^2 + 5^2}} = \frac{v_1}{\sqrt{26}} = \begin{bmatrix} \frac{1}{\sqrt{26}} & 0 \\ 0 & \frac{5}{\sqrt{26}} \end{bmatrix}$$

$$n_1 = \begin{bmatrix} \frac{1}{\sqrt{26}} & 0 \\ 0 & \frac{5}{\sqrt{26}} \end{bmatrix}$$

$$v_2 = o_2 - \frac{\langle o_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = \begin{bmatrix} \frac{25}{13} & 0 \\ 0 & \frac{-5}{13} \end{bmatrix}$$

$$n_2 = \begin{bmatrix} \frac{5}{\sqrt{26}} & 0 \\ 0 & \frac{-1}{\sqrt{26}} \end{bmatrix}$$

$$v_3 = o_3 - \frac{\langle o_3, v_1\rangle}{\langle v_1, v_1\rangle} v_1 - \frac{\langle o_3, v_2\rangle}{\langle v_2, v_2\rangle} v_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

$$n_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

$$v_4 = o_4 - \frac{\langle o_4, v_1\rangle}{\langle v_1, v_1\rangle} v_1 - \frac{\langle o_4, v_2\rangle}{\langle v_2, v_2\rangle} v_2 - \frac{\langle o_4, v_3\rangle}{\langle v_3, v_3\rangle} v_3 = \begin{bmatrix} 0 & 6 \\ 0 & 0 \end{bmatrix}$$

$$n_4 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

n_1, n_2, n_3, n_4 form an orthogonal basis for S. Hence,

$$S = [\begin{bmatrix} \frac{1}{\sqrt{26}} & 0 \\ 0 & \frac{5}{\sqrt{26}} \end{bmatrix}, \begin{bmatrix} \frac{5}{\sqrt{26}} & 0 \\ 0 & \frac{-1}{\sqrt{26}} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}]$$

2. *PageRank*(**#transformations**, **#computationaltools**)

   In this problem, we will explore a toy model of the Internet and Google's search algorithm, Page Rank. In our toy Internet, there will be a total of 6 webpages, denoted as nodes on a graph, as shown in Fig. 4 below.

   Each edge will correspond to a link from one page to another. Note that pages that link back to one another are represented by double arrows (but you may think of them as two different edges). The main idea of the Page Rank algorithm is to find which webpages are "better" by taking into account both *the number of links* and *the quality of these links.*

   (a) If Mark is currently on a webpage, then we will assume that after some time he will click on one of the links and end up on a different webpage. If a page has multiple links, there is an equal probability that he will click on any of those links. For instance, a user on webpage 1 has a 50%-50% chance of moving either to page 2 or 4 (the two links leaving page 1). We can create a Markov matrix to model this situation: consider that the columns correspond to Mark's current webpage, and the rows to the next webpage he reaches. Write a Markov matrix M that models the probabilities of moving from one page to another on our toy Internet.

   Answer:

   Each of the columns in the matrix should all add to 1. Each value in the matrix represents the probability that Mark will visit a website next, given that he is in the position provided by the vector being multiplied by the matrix.

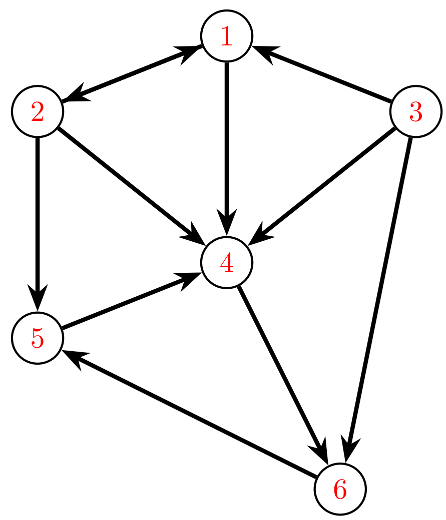

Figure 4: Our toy Internet

Assuming that Mark is equally likely to click on any link, the probability that he will visit some webpage linked to the site he is currently in is 1/n, where n is the total number of websites linked to that site. However, it is zero if the two websites are not linked. Therefore, the probability matrix M is the following:

$$M = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{3} & 0 & 1 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{1}{3} & 1 & 0 & 0 \end{bmatrix}$$

(b) Assume that Mark can begin his Internet adventure on any of these six webpages, with equal probability. Create a vector u_0 that represents this initial probability distribution. Compute the probability that Mark will be on each of the 6 pages after

i. 1 time step;
ii. 2 time steps;
iii. 3 time steps;
iv. 10 time steps.

Answer:

Assuming that Mark has an equal probability of starting in any webpage, then

$$\vec{u}_0 = \begin{pmatrix} 0.167 \\ 0.167 \\ 0.167 \\ 0.167 \\ 0.167 \\ 0.167 \end{pmatrix}$$

Therefore, $\quad \vec{u}_n = M^n \vec{u}_0$.

$$\vec{u}_1 = \begin{pmatrix} 0.1111 \\ 0.0835 \\ 0 \\ 0.3617 \\ 0.2226 \\ 0.2226 \end{pmatrix}; \quad \vec{u}_2 = \begin{pmatrix} 0.0278 \\ 0.0556 \\ 0 \\ 0.3060 \\ 0.2504 \\ 0.3617 \end{pmatrix}; \quad \vec{u}_3 = \begin{pmatrix} 0.0185 \\ 0.0139 \\ 0 \\ 0.2828 \\ 0.3802 \\ 0.3060 \end{pmatrix}; \quad \vec{u}_{10} = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 0.3959 \\ 0.3169 \\ 0.2885 \end{pmatrix}$$

(c) Do your findings in part (b) match your expectations? Use the graph structure and the eigenvectors of matrix M to justify.

Answer:

Yes, it matches my expectations. I assumed that the system would eventually reach out a stationary distribution, (i.e., a steady state equilibrium). My observation from Figure 4 is that multiple sites redirected to page 4, and its only connection is to page 6, and in turn its only connection is to page 5. This means that all of the traffic is redirected to the loop in pages 4, 6, and 5 but they cannot come back to pages 1, 2, or 3. Thus I expected Mark to have a higher probability of landing in either of these three nodes and that in the end the equilibrium state would be close to the vector [6]

$$\vec{V} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$$

(d) What if we knew for sure that Mark started browsing on webpage 4? Create a new initial vector v_0 that represents this initial probability distribution. Compute the probability that Mark will be on each of the pages after

i. 1 time step;

[6] networks: The student identified and explained some of the network dynamics. For example, once you enter the loop 6-5-4-6, it is impossible to exit through a simple redirect and thus all traffic will eventually make it to the loop and thus the likelihood of finishing in either of these spots is close to 1/3

ii. 2 time steps;

iii. 3 time steps.

Answer:

The vector that would represent the starting position of Mark in webpage 4 is the following:

$$\vec{u}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

Moreover, the following states are:

$$\vec{u}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}; \quad \vec{u}_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}; \quad \vec{u}_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

(e) How do your findings compare with those in part (b)? Use the graph structure to justify.

Answer:

The equilibrium here is different, because it consists of a closed loop. Nodes can get into the system, but they cannot exit. This means that all of the traffic in one node moves to the next, and to the next, until it completes yet another loop. All of this in a perpetual cycle.

(f) As you probably know, Internet browsing is not as deterministic as our toy model. Let's consider another extreme: completely random browsing. That is, after one time step, Mark has an equal chance of being on any of the other five pages. Write a Markov matrix R that models the probabilities of moving from one page to another in this completely random browsing mode. Describe the expected long-term probability distribution of random browsing, using the eigenvectors of matrix R to justify.

Answer:

Assuming that Mark has an equal chance of moving to any website, then the Markov matrix representing all of this is the following:

$$R = \begin{bmatrix} 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 \end{bmatrix}$$

The eigenvector to the eigenvalue 1 of a Markov matrix is itself called the stable equilibrium distribution, and in this case it is the vector

$$\vec{u}_0 = \begin{pmatrix} \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \end{pmatrix};$$

This is simple to visualize. If the browsing is fully random, and every webpage is equally connected, then every website should have an equal probability of being landed on. Moreover, without loss of generality every webpage would be symmetrical in that they would all be equal to each other (for example, each would be connected to all the five other webpages).

(g) Finally, consider a mix of deterministic and random browsing. Consider a matrix A which is a weighted sum of the two Markov matrices: $A = \frac{3}{4}M + \frac{1}{4}R$. Verify that A is a Markov matrix and then describe the expected long-term probability distribution of mixed browsing, using the eigenvectors of matrix A to justify.

Answer:

The new matrix would also be a Markov matrix. This is simple to visualize because the sum of every column equals 1, and thus if the matrices M and R are weighted by 3/4 and 1/4 respectively, then every column vector will add up to 3/4 and similarly every column vector of R will add up to 1/4 for a total of 1. The new matrix would be:

$$A = \begin{bmatrix} 0 & \frac{3}{10} & \frac{3}{10} & \frac{1}{20} & \frac{1}{20} & \frac{1}{20} \\ \frac{17}{40} & 0 & \frac{1}{20} & \frac{1}{20} & \frac{1}{20} & \frac{1}{20} \\ \frac{1}{20} & \frac{1}{20} & 0 & \frac{1}{20} & \frac{1}{20} & \frac{1}{20} \\ \frac{17}{40} & \frac{3}{10} & \frac{3}{10} & 0 & \frac{4}{5} & \frac{1}{20} \\ \frac{1}{20} & \frac{3}{10} & \frac{1}{20} & \frac{1}{20} & 0 & \frac{4}{5} \\ \frac{1}{20} & \frac{1}{20} & \frac{3}{10} & \frac{4}{5} & \frac{1}{20} & 0 \end{bmatrix}$$

You can see that if we add the columns, they all add to 1 as predicted.

$$\vec{u}_1 = \begin{pmatrix} 0.1253 \\ 0.1044 \\ 0.0418 \\ 0.3131 \\ 0.2088 \\ 0.2088 \end{pmatrix}; \quad \vec{u}_2 = \begin{pmatrix} 0.0804 \\ 0.0919 \\ 0.0480 \\ 0.2745 \\ 0.2223 \\ 0.2849 \end{pmatrix}; \quad \vec{u}_3 = \begin{pmatrix} 0.0810 \\ 0.0756 \\ 0.0477 \\ 0.2682 \\ 0.2757 \\ 0.2537 \end{pmatrix}; \quad \vec{u}_{10} = \begin{pmatrix} 0.0770 \\ 0.0752 \\ 0.0477 \\ 0.2897 \\ 0.2507 \\ 0.2617 \end{pmatrix}$$

The eigenvector of A for the eigenvalue 1 is the following:

$$\vec{u} = \begin{pmatrix} 0.0768 \\ 0.0751 \\ 0.0476 \\ 0.2850 \\ 0.2530 \\ 0.2625 \end{pmatrix}$$

As we can observe, the solution tends towards the equilibrium (i.e., the eigenvector for the eigenvalue 1) but it hasn't yet reached perfect equilibrium.

(h) The probability of landing on a given page at the end corresponds to its Page Rank. Given our mixed model (part (g)), in what order would our search engine present these pages?

Answer:

In the equilibrium, $A\vec{u}_0 = \vec{u}_0$. Therefore, $(A - I)\vec{v} = \vec{0}$ and we solve for $\vec{v}$ then we will find that the engine would order these pages as follows: 4,6,5,1,2,3. [7]

[7] algorithm: The students showed mastery of the Page Rank algorithm throughout this entire problem. For example, the student explained important concepts like the fact that the eigenvector for the eigenvalue 1 is the stationary distribution. Moreover, the student attempted to help the reader visualize why under perfect random behaviour you will have an equal chance of ending in any one webpage, and why under our current example you will enter a loop that you cannot exit.

3. *Animation Studios*(**#transformations**, **#computationaltools**) In class, we explored how linear transformations can be used in computer graphics to rotate, scale, reflect and project graphic objects. One operation that we still do not know how to perform is translation, that is, moving an object up/down, left/right.

   (a) Describe why linear transformations cannot perform translations. (Hint: Consider the zero-vector)

   Answer:

   Linear transformations need to satisfy the following two conditions: $T(\vec{v}_1+\vec{v}_2) = T(\vec{v}_1) + T(\vec{v}_2)$ and $T(k\vec{v}) = kT(\vec{v})$. Therefore, they cannot perform translations or else they would not satisfy these conditions for the zero-vector.

   We thus introduce the idea of *affine transformations*. Those are transformations of the type $T(\vec{v}) = A\vec{v} + \vec{b}$.

   (b) What would the matrix A and the vector $\vec{b}$ be to perform a translation of one unit to the right? Test your affine transformation on the vector $v = \langle 2, 1\rangle$

   Answer:

   The matrix A would be equal to the identity matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, and the vector b would be equal to the unit vector pointing to the right $\vec{b} =< 1, 0 >$.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} 2, & 1 \end{pmatrix} + \begin{pmatrix} 1, & 0 \end{pmatrix} = \begin{pmatrix} 3, & 1 \end{pmatrix}$$

   (c) We may combine transformations, for example, by rotating a vector by 90 degrees counterclockwise, and then translating it up by one unit as a single affine transformation. What should A and $\vec{b}$ be to perform these two transformations? Again, test it with $v = \langle 2, 1\rangle$.

   Answer:

   The matrix A should mimic the following notation $\begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$ thus, the matrix A equals $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Moreover, the vector b should be equal to the unit vector pointing up $\vec{b} =< 0, 1 >$.

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} 2, & 1 \end{pmatrix} + \begin{pmatrix} 0, & 1 \end{pmatrix} = \begin{pmatrix} -1, & 3 \end{pmatrix}$$

   (d) *Order matters!* What if we had instead translated the vector $\vec{v}$ one unit up first, and then rotated it by 90 degrees counterclockwise? What would be the matrix A and vector $\vec{b}$ that perform these operations, written in the standard affine form $A\vec{v} + \vec{b}$?

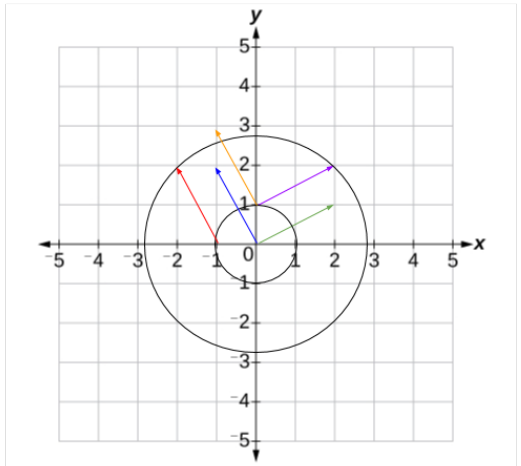

Figure 5: Rotation and translation. The order of operations matter.

Answer:
The order of operations matter. The original vector (green arrow) is rotated 90º counterclockwise (blue arrow) and then it is finally translated one unit north (orange arrow) in part (c). However, now we are asked to translate the vector up one unit (purple arrow) and then rotate 90º counterclockwise (red arrow).
Moreover, the matrix can be obtained by translating back to the original origin, doing the rotation, and returning to the previous coordinate system as explained in the forum math stack exchange. [8]
Therefore, the matrix A is the following:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

And the vector b is the zero-vector: $\vec{b} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

Hence the full equation is: $\begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}$

[8] https://math.stackexchange.com/questions/2093314/rotation-matrix-of-rotation-around-a-point-other-than-the-origin

Technically, figure 5 is a graphic solution and we can observe that the results from the matrix multiplication and the graphic agree with each other.

(e) Compare your matrices A and $\vec{b}$, and your final transformed vector $A\vec{v}+\vec{b}$ from parts (c) and (d).

Answer:

As can be observed from figure 5, the resulting vectors (orange arrow for part (c) and red arrow for part (d)) are different. The order of operations matter because the rotations happen with respect to the origin so that everything remains equally distant before and after the rotation, while this is not true for the translation.

(f) Finally, we can "cheat" to encode everything as a single matrix multiplication. We construct a combined vector V given as:

$$V = \begin{bmatrix} v \\ 1 \end{bmatrix}$$

and a combined matrix M:

$$M = \begin{bmatrix} A & \vec{b} \\ 0 \cdots 0 & 1 \end{bmatrix}$$

Describe the result of multiplying the matrix M by the vector V.

Answer:

Oops. I believe that's what I did above. The result of multiplying the matrix times the vector is the same as doing all of the long procedure. The reason that this happens is because the matrix encompasses all of the same information in a more compact and data-dense form.

(g) Construct the combined vector V and matrix M that perform each of the affine transformations in parts (b) - (e) and verify your results.

Answer:

Redo of part (b):

According to the cheat sheet, the vector V would be $\vec{V} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ and the combined matrix M would be $M = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$, therefore, $M\vec{V} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$

We can observe that the answer of $\begin{pmatrix} 3, & 1 \end{pmatrix}$ is the same.

Redo of part (c):

According to the cheat sheet, the vector V would be $\vec{V} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ and the combined matrix M would be $M = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$, therefore, $M\vec{V} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ 2 \end{pmatrix}$
We can observe that the answer of $\begin{pmatrix} -1, & 3 \end{pmatrix}$ is the same.

1 Appendix

R-Code:
https://gist.github.com/almagashi/baeee27a8f7c61b445e8afdc15e7f4ff
SageMath:
https://gist.github.com/almagashi/7daa995748fcf7f55d36223816649134