DATA 1010
Problem Set 1
Due 14 September 2018 at 11 PM

Problem 1

Consider a table with 3 columns and 1000 rows, some of whose entries are missing. Denote by A the set of rows with an entry in the first column, B the set of rows with an entry in the second column, and C the set of rows with an entry in the third column. Use set notation (intersections, unions, and complements) to represent the following sets in terms of A, B, and C.

(i) The set of rows with no missing entries

(ii) The set of rows with all missing entries

(iii) The set of rows with at least one entry present

(iv) The set of rows with an entry in the first column and exactly one other entry

Solution

(i) A row has no missing entries if it's in A and B and C. Therefore, the answer is $A \cap B \cap C$.

(ii) Similarly, a row has all of its entries missing if it's in A^{c} and B^{c} and C^{c}. So the answer is $A^{\mathsf{c}} \cap B^{\mathsf{c}} \cap C^{\mathsf{c}}$.

(iii) At least one entry present means that it is *not* the case that all of the entries are missing, so $(A^{\mathsf{c}} \cap B^{\mathsf{c}} \cap C^{\mathsf{c}})^{\mathsf{c}}$ is the right set.

(iv) To satisfy this condition a row must be in A and either in $B \cap C^{\mathsf{c}}$ or $B^{\mathsf{c}} \cap C$. So the answer is

$$A \cap ((B \cap C^{\mathsf{c}}) \cup (B^{\mathsf{c}} \cap C))$$

Problem 2

Implement the matrix multiplication algorithm from scratch in Julia (that is, any multiplication operations used must be multiplications of two *numbers*). Check your function using the following line:

```
julia> myprod([2 3 4; -4 2 5],[1 2 -4; -6 5 2; 0 1 0 ])
2×3 Array{Int64,2}:
-16  23  -2
-16   7  20
```

Solution

We begin by writing a function for calculating the dot product, and then we loop through the entries of the product matrices and compute them one at a time:

```
function mydot(a,b)
    s = 0.0
    for i = 1:length(a)
        s += a[i] * b[i]
    end
    s
end
```

```
function myprod(A,B)
    m = size(A,1)
    n = size(B,2)
    C = zeros(m,n)
    for i=1:m
        for j=1:n
            C[i,j] = mydot(A[i,:],B[:,j])
        end
    end
    C
end
```

Problem 3

Suppose that U and V are vector subspaces of $\mathbb{R}^n$. Show that $U \cup V$ is *not* a subspace of $\mathbb{R}^n$ unless $U \subset V$ or $V \subset U$.

Solution

Suppose that U and V are vector subspaces of $\mathbb{R}^n$ and that we do not have $U \subset V$ or $V \subset U$. Then there exists a vector $\mathbf{u} \in U$ which is not in V and a vector $\mathbf{v} \in V$ which is not in U. We claim that $\mathbf{u} + \mathbf{v}$ is not in U or V.

If $\mathbf{u} + \mathbf{v}$ were in U, then $\mathbf{u} + \mathbf{v} + (-\mathbf{u}) = \mathbf{v}$ would be in U as well, since U is closed under vector addition. But we know that $\mathbf{v}$ is not in U. Similarly, $\mathbf{u} + \mathbf{v}$ is also not in V.

Since $\mathbf{u}$ and $\mathbf{v}$ are both in $U \cup V$ but their sum is not in $U \cup V$, we conclude that $U \cup V$ is not a subspace.

Problem 4

Show that if $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent and $\{\mathbf{v}_1 + \mathbf{w}, \mathbf{v}_2 \ldots, \mathbf{v}_n\}$ is linearly dependent, then $\mathbf{w}$ is in the span of $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.

Solution

If $\{\mathbf{v}_1 + \mathbf{w}, \mathbf{v}_2 \ldots, \mathbf{v}_n\}$ is linearly dependent, then there exists a nontrivial linear combination of these vectors which is equal to the zero vector, say

$$c_1(\mathbf{v}_1 + \mathbf{w}) + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}.$$

If c_1 were zero in this equation, then there would be a nontrivial vanishing linear combination of the vectors $\mathbf{v}_2, \ldots, \mathbf{v}_n$, which isn't possible since we know those vectors are linearly independent. Therefore, we can solve for $\mathbf{w}$ to get

$$\mathbf{w} = -\frac{c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n}{c_1}.$$

Therefore, $\mathbf{w}$ is in the span of $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.

Problem 5

If A is a full-rank $m \times n$ matrix, then the vector in the span of the columns of A which is closest to $\mathbf{b} \in \mathbb{R}^m$ is $A\widehat{\mathbf{x}}$, where $\widehat{\mathbf{x}} = (A'A)^{-1}A'\mathbf{b}$.

Find the vector in the span of the columns of A which is closest to $\mathbf{b} = [4, 2, -1]$, where

$$A = \begin{bmatrix} 1 & -2 & 4 \\ -3 & -1 & 2 \\ 0 & 1 & -2 \end{bmatrix}$$

Solution

The first two columns of A are linearly independent, but the third column is in the span of the first two. Therefore, the span of the columns of A is the same as the span of the columns of

$$B = \begin{bmatrix} 1 & -2 \\ -3 & -1 \\ 0 & 1 \end{bmatrix}$$

So we can find the vector in the range of A which closest to $\mathbf{b}$ by finding the vector in the range of B which is closest to $\mathbf{b}$:

$$\widehat{\mathbf{b}} = (B'B)^{-1}B'\mathbf{b} = [-1/59, -108/59] \approx [0.017, -1.83],$$

using `(B' * B) \ (B' * [4,2,-1])`.

Problem 6

Suppose that $A = U\Sigma V'$ where Σ is diagonal and U and V are orthogonal matrices. Show that the columns of U are eigenvectors of AA' and that the columns of V are eigenvectors of $A'A$.

Hint: substitute $A = U\Sigma V'$ into the expressions AA' and $A'A$.

Solution

We have

$$AA' = U\Sigma V'V\Sigma U' = U\Sigma^2 U'.$$

Multiplying on the right by U, we have that

$$AA'U = U\Sigma^2.$$

The jth column on the left is the product of AA' with the jth column of U, while the jth column on the right is σ_j^2 times the jth column of U, where σ_j is the jth diagonal entry of Σ. Therefore, the jth column of $\mathbf{u}$ is an eigenvector of A.

Similarly, we have

$$A'A = V\Sigma^2 V',$$

and the same argument shows that the columns of V are all eigenvectors of $A'A$.

Problem 7

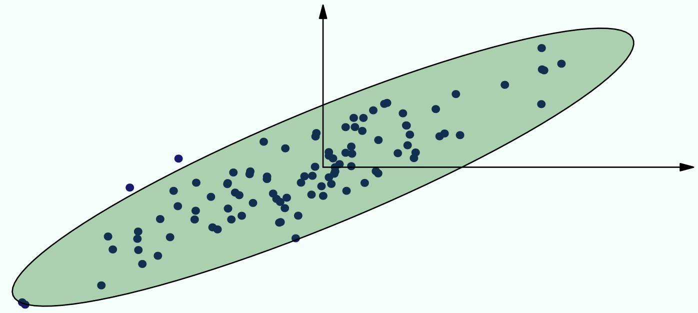
The singular value decomposition can be used to identify the primary axes in a ellipsoidal point cloud. Run the following block to generate and plot a set of 100 points.

```julia
using LinearAlgebra
using Plots
numpoints = 100
T = [1 2; 0 1]
P = (T * randn(2,numpoints))'
scatter(P[:,1],P[:,2],aspect_ratio=:equal)
```

Note that the coordinates of the points are stored in the rows of P.

Use Julia to compute the singular value decomposition $U\Sigma V'$ of P, and show visually that the columns of V run along the axes of the ellipse that fits the point cloud (the one shown in the figure).

Hint: `plot!([(a,b),(c,d)])` adds a line segment from the point (a, b) to the point (c, d) to the current plot. You'll want to plot line segments representing both of the columns of V.



Solution

In addition to the code block above, we use

```
U, Σ, V = svd(P)
plot!([(0,0),(V[1,1],V[2,1])])
plot!([(0,0),(V[1,2],V[2,2])])
```

to add the given vectors to the figure. We see that these vectors do indeed appear to run along the axes of the elliptical point cloud.