CS 189 HW0

Huanjie Sheng (25928718)

January 18, 2018

1 Getting Started

(a) Who else did you work with on this homework? In case of course events, just describe the group. How did you work on this homework? Any comments about the homework?

> I worked on this homework with Weiran Liu ($weiran_{l}iu@berkeley.edu$).
>
> First, I downloaded MikTeX and TeXstudio and spent about one hour learning how to use LaTeX.
>
> Second, I did the work using my brain.
>
> Third, I did the homework using Matlab on my laptop and looked up the questions online.
>
> Fourth, I checked my answers in step 2 by comparing results obtained using in the step 3.
>
> Last, I submitted on Piazza.
>
> I think we should be provided with a template of the homework rather than just a PDF file.

(b) Please copy the following statement and sign next to it. We just want to make it extra clear so that no one inadverdently cheats.

> *I certify that all solutions are entirely in my words and that I have not looked at another students solutions. I have credited all external sources in this write up.*
>
> Signature: 盛焕杰

2 Sample Submission

submission.txt

$$
\begin{array}{c}
1 \\
1 \\
1 \\
1 \\
1
\end{array}
$$

3 Linear Algebra Review

Calculate the matrix M first:

$$M = uv^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} * \begin{bmatrix} 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 \times 2 & 1 \times 3 \\ 2 \times 2 & 2 \times 3 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$$

(a) Compute the eigenvalues and eigenvectors of the matrix M.

Here I denote eigenvectors are x, and the corresponding eigenvalues are λ.
By definition, the eigenvalues and eigenvectors of the matrix M satisfy the following equation:

$$Mx = x\lambda \tag{1}$$

Therefore, we can solve the eigenvalues of M by solving the above linear equation, which is equivalent to find the root of the following equation:

$$\left| \lambda I - M \right| = 0$$

That is,

$$\begin{vmatrix} \lambda - 2 & -3 \\ -4 & \lambda - 6 \end{vmatrix} = 0$$

$$(\lambda - 2)(\lambda - 6) - (-3)(-4) = 0$$

$$\lambda(\lambda - 8) = 0$$

$$\lambda_1 = 0 \text{ or } \lambda_2 = 8$$

Substitute the λ values back into equation (1), and we can solve for the non-trivial solutions as eigenvectors x. For example, for $\lambda = 0$, we have:

$$\begin{bmatrix} -2 & -3 \\ -4 & -6 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} -2 & -3 & 0 \\ -4 & -6 & 0 \end{array}\right] \Rightarrow \left[\begin{array}{cc|c} 2 & 3 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

As a result, the eigenvectors are:

$$x_{\lambda_1} = \begin{bmatrix} 3s_1 \\ -2s_1 \end{bmatrix}, x_{\lambda_2} = \begin{bmatrix} s_2 \\ 2s_2 \end{bmatrix} \text{, where } s_1, s_2 \in \mathbb{R}$$

Especially, the normalized eigenvectors are

$$x_{\lambda_1} = \begin{bmatrix} \frac{3}{\sqrt{13}} \\ -\frac{2}{\sqrt{13}} \end{bmatrix} \text{ or } \begin{bmatrix} -\frac{3}{\sqrt{13}} \\ \frac{2}{\sqrt{13}} \end{bmatrix}, x_{\lambda_2} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} \text{ or } \begin{bmatrix} -\frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix}$$

(b) Compute the rank and the determinant of the matrix M. What is the dimension of the nullspace of the matrix M?

We can compute the rank of M by converting the matrix to its Echelon form:

$$M = \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & 3 \\ 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \frac{3}{2} \\ 0 & 0 \end{bmatrix}$$

There is only one row containing non-zero elements, so

$$\text{rank}(M) = 1$$

By definition, the determinant of the matrix M can be calculated as following:

$$det(M) = \begin{vmatrix} 2 & 3 \\ 4 & 6 \end{vmatrix} = 2 \times 6 - 4 \times 3 = 0$$

By definition, the nullspace of the matrix M is the solutions to the following linear equation:

$$Mx = 0$$

That is,

$$\begin{bmatrix} 1 & \frac{3}{2} \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + \frac{3}{2}x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$x_1 = -\frac{3}{2}x_2$$

$$x = \begin{bmatrix} -\frac{3}{2}s \\ s \end{bmatrix} = \begin{bmatrix} -\frac{3}{\sqrt{13}}s \\ \frac{2}{\sqrt{13}}s \end{bmatrix} \text{, where } s_1, s_2 \in \mathbb{R}$$

The nullspace of the matrix M is a line and the dimension is 1. A more direct solution is to look at rank(M), and uses rank-nullity theorem.

(c) Now consider two non-zero vectors p and q in $\mathbb{R}^d$ and the matrix $N = pq^T$. Repeat the computations for parts (a) and (b) for the matrix N.

First of all, the order of this question is extremely misleading. I will have to prove that rank(N) = 1 before getting the actual eigenvalues.

(c.b) According to the wikipedia page of rank, for given matrix A in $\mathbb{R}^{m\times n}$ and matrix B in $\mathbb{R}^{n\times k}$, we have $rank(AB) <= min(rank(A), rank(B))$. Therefore, $rank(N) <= min(rank(p), rank(q^T)) = 1$. Because both p and q are non-zero vectors, N must be non-zero too. Thus, $0 < rank(N) <= 1$. All in all, we have $rank(N) = 1$.

Secondly, according to rank-nullity theorem on wikipedia, we have $rank(N) + null(N) = d$. it's obvious that the dimension of nullspace is d-1.

Because $rank(N) = 1$, according to the properties of determinant on wikipedia, $det(N) = 0$.

(c.a) Thanks to ***Theo Bendit*** on stackexchange, we know that the nullspace of the matrix N is just the eigenspace corresponding to eigenvalue 0. That means, we have $d-1$ eigenvalues that all equal 0.

Now, I'm hoping to prove that the vector p is actually a eigenvector of the matrix N, and the corresponding eigenvalue is q^Tp. It is obvious because q^Tp is a scalar:

$$pq^Tp = q^Tpp$$

We are left with one thing that is to find the eigenvectors of those zero eigenvalues aka the null space. My guess would be:

$$x_{\lambda_2} = \begin{bmatrix} -(d-1)q_2q_3...q_d \\ q_1q_3...q_d \\ . \\ . \\ . \\ q_1q_2...q_{d-1} \end{bmatrix}, x_{\lambda_i} = \begin{bmatrix} q_2q_3...q_d \\ q_1q_3...q_d \\ . \\ -(d-1)q_1q_2..q_{i-2}q_{i-1}..q_d \\ . \\ q_1q_2...q_{d-1} \end{bmatrix}, x_{\lambda_d} = \begin{bmatrix} q_2q_3...q_d \\ q_1q_3...q_d \\ . \\ . \\ -(d-1)q_1q_2...q_{d-2}q_d \\ q_1q_2...q_{d-1} \end{bmatrix}$$

I will just prove x_{λ_2} because the rest are the same. This is done by simply writing out pq^T and x_{λ_2}:

$$pq^Tx_{\lambda_2} = \begin{bmatrix} p_1q_1 & p_1q_2 & ... & p_1q_d \\ p_2q_1 & p_2q_2 & . & p_1q_d \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ p_dq_1 & p_dq_2 & ... & p_dq_d \end{bmatrix} \begin{bmatrix} -(d-1)q_2q_3...q_d \\ q_1q_3...q_d \\ . \\ . \\ . \\ q_1q_2...q_{d-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix}$$

Besides, it is not difficult to understand that eigenspace for $\lambda = 0$ is just the nullspace, and

$$\begin{bmatrix} q_2q_3...q_d \\ q_1q_3...q_d \\ . \\ . \\ . \\ -(d-1)q_1q_2...q_{d-1} \end{bmatrix}$$ is a linear combination of $x_{\lambda_2}...x_{\lambda_d}$ because the dimension of nullspace is just d-1

4 Linear Regression and Adversarial Noise

(a) Can the adversary always fool us by seting a particular value for exactly one ϵ_i

No, the adversary cannot fool us.

Because x_i's are all distinct and non-zero. We can write the solution to the standard ordinary least-squares regression as $w = (A^T A)^{-1} A^T y$. Here, A represents the matrix $[x, 1]$ Another way to look at this, as is taught by Professor Jonathan Shewchuk, is to consider $P = (A^T A)^{-1} A^T$ the pseudo inverse matrix of x. Thus, the solution is just a linear transformation of y from $\mathbb{R}^n$ to $\mathbb{R}^2$. A shortcut would be that we know w has two degrees of freedom so manipulating one observation is not enough to get the whole space. For a more thorough counter example, we could assume y_1 is being manipulated and consider:

$$w = Py = \begin{bmatrix} p_{11} & p' \\ p_{21} & p'' \end{bmatrix} \begin{bmatrix} y_1 \\ y' \end{bmatrix} = \begin{bmatrix} p_{11}y_1 + p'y' \\ p_{21}y_1 + p''y' \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

We can cancel y_1 and get the linear relationship between w_1 and w_2

$$p_{21}w_1 - p_{11}w_2 = p_{11}p''y' - p_{21}p''y'$$

Therefore, setting one observation will only screw up our solution to be a line (hyperplane) not the entire space. A simple counterexample would be two points, one is the origin and the other is (x_1, y_1), where $x_1 \neq 0, x_2 \neq 0$, it's not possible to get $a = \frac{x_2}{x_1}, b = x_2$

(b) How about setting two observations?

Yes. In this case, the adversary can fool us.

We can perform a similar matrix calculation to what we did before assuming the observations we are going to manipulate are y_1 and y_2:

$$w = Py = \begin{bmatrix} p_{11} & p_{12} & p' \\ p_{11} & p_{22} & p'' \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y' \end{bmatrix} = \begin{bmatrix} p_{11}y_1 + p_{12}y_2 + p'y' \\ p_{21}y_1 + p_{22}y_2 + p''y' \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

It's not possible to cancel both y_1 and y_2 here. Therefore, $w_1 \in \mathbb{R}$ and $w_2 \in \mathbb{R}$. **However, if we know what adversary are going to do on our data, we can compensate for it!** Well, this is simple and trivial.

(c) In the context of machine learning and applications, what lessons do you take-away after working through this problem?

I learned several points:

1. Noise is scary
2. Adversary is dangerous
3. Life is hard
4. **(Seriously) the standard ordinary least square is generally not robust and can be readily screwed by outliers.**

5 Background Review

Please describe the coursework that you have undertaken on the following topics:

(a) Linear Algebra, e.g., EE 16A/B

I took linear algebra about eight years ago and can still remember some stuff but not everything.

(b) Optimization, e.g., EECS 127

I never had any classes in optimization algorithms but I did had some practical experience with evolution dynamics in my previous research.

(c) Probability and stochastic processes, e.g., EECS 126

I took three different courses in probability and stochastic processes and I almost use the knowledge in all of my research projects.

(d) Vector Calculus, e.g., EECS 127, Math 53.

I cannot remember anything about vector calculus even if I've learned it before. However, it seems to be very intuitive.

I have zero experience in python but I'm very familiar with C/C++, Java, Matlab, Mathematica. I had very strong fundamental knowledge about programming. I hope to learn python but there was never a motivation to do so. This could be a great chance. By the way, I'm old-fashion so I don't like scripting programming languages.

6 Your Own Question

Big question: Can machine learns how to choose models?
This question is too broad and should be considered as unresolvable based on my knowledge. Therefore, it's a bad question.

Subquestion: Given a series of data points $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, how do I find out the best analytical fitting solution?
I cannot find a complete mathematical solutions to this problem, but I'm find a metric called Akaike information criteria (AICc). It was calculated by the formula

$$AICc = n\dot{l}og_e(RSS/n) + 2K + 2K(K+1)/(n-K-1),$$

where n is the number of samples, RSS is the residual sum of squares (or sum of squared error), and K is the number of parameters in the model. The difference in the AICc values (ΔAICc) was used to evaluate the relative strength of different models.

This partially addressed my question but it cannot prove that the model with the lowest AICc is the best model, because it's not possible to calculate AICc for all analytical models.

This is a very important question because most of the time we don't know the mechanisms of the system when we analyze the data from the experiments. Using simple functions like linear functions to fit the data can only gives you insight into the first-order component of the taylor series. That's usually not enough for complex systems like biological systems. I will keep looking for solutions. Meanwhile, because my lack of knowledge, it's possible that this is a simple problem and there is a thorough mathematical solution to it. I'm happy to learn that.