Alternating-direction-method of Multipliers-Based Adaptive Nonnegative Latent Factor Analysis: Supplementary Materials

Yurong Zhong, Shangce Gao, *Senior Member*, *IEEE*, and Xin Luo, *Senior* *Member*, *IEEE*

This is the supplementary file for the paper entitled *Alternating-direction-method of Multipliers-Based Adaptive Nonnegative Latent Factor Analysis.* Detailed convergence proof of A2NLF and additional figures are presented here.

# Convergence of A2NLF

1. *Proof of Lemma* 1

Note that learning objective of A2NLF-PSO/A2NLF-TPE is non-convex. According to [48], any of its limit points where the gradient becomes zero can be a local/global optimum, or saddle point. Hence, such a limit point can be treated as a solution. Supposing that the optimal solution to *au*,*k*(*q*) by (4c) is *at* *u*,*k*(*q*). Thus, it fulfills the following condition:

 (S1)

Following (4e) and (5c), by applying the update rule of *hu*,*k*(*q*) to (S1), we have:

 (S2)

Then (22a) stands based on (S2). Following the same principle, we can derive the optimality condition of (5*b*) related to *xi,k*(*q*):

 (S3)

Then (22b) holds based on (S3). Hence, *Lemma* 1 holds, and **Step 1** is implemented.□

1. *Proof of Lemma* 2

Considering the difference between  and , we have:

 (S4)

where (\*) performs the second-order Taylor expansion of the left term. Then, considering (5a)’s optimality condition, (S4) is transformed as:

 (S5)

Thus, the difference between and  is:

 (S6)

Moreover,  and  yields:

 (S7)

where (I) is based on the update rules of (*hu*,*k*(*q*), *wi*,*k*(*q*)) given in (4e), (4f) and (5c), and (II) is achieved with *Lemma* 1.With (S5)-(S7), we have the following deduction:

 (S8)

Owing to (24a), (25a) stands, which indicates that the augmented Lagrangian function (3) related to *s*(*q*) or *τq* is non-increasing as *at* *u*,*k*(*q*)>0 and *xt* *i*,*k*(*q*)>0. Then after the *t-*th iteration, the partial objective from (3) related to *s*(*q*) or *τq* is formulated as:

  
(S9)

By substituting (S2) and (S3) into (S9), we have:

 (S10)

(S10) indicates that if (24b) is fulfilled, (25b) holds, thereby making (3) related to *s*(*q*) or *τq* lower-bounded as *at* *u*,*k*(*q*)>0 and *xt* *i*,*k*(*q*)>0. Based on the above inferences, *Lemma* 2 stands, and **Step 2** is implemented.□

1. *Proof of Theorem* 1

***Part a*.** Following *Lemma* 2,  converges as *t*→∞, indicating that:

 (S11)

With (24), when (S1) is fulfilled, the upper-bound of  is zero as *t*→∞, thereby achieving (26). Following (S8) and (26), we have [24]:

 (S12a)

 (S12b)

 (S12c)

 (S12d)

 (S12e)

 (S12f)

Based on (22) and (S12), we have the following inferences:

 (S13a)

 (S13b)

Based on (4e), (4f) and (S13), we conclude that (27) is fulfilled.

***Part b*.** Firstly, following (4a), (4b) and (5a), the update rules of (*pu*,*k*(*q*), *zi*,*k*(*q*)) can be rearranged as:

 (S14a)

 (S14b)

Then by substituting (27) and (S12) into (S14), we have:

 (S15a)

 (S15b)

Hence, considering a limit point  of a sequence  generated by the update rules of {*ψ*1(*q*), *ψ*2(*q*)} based on (4) and (5), the following KKT conditions are satisfied with (27) and (S15):

 (S16a)

 (S16b)

 (S16c)

 (S16d)

Afterwards, considering the remaining KKT conditions regarding constraints *au,k*(*q*)>0 and *xi,k*(*q*)>0, we extend the original augmented Lagrangian *g*:

 (S17)

where the operator *Tr*(⋅) computes the trace of an enclosed matrix, and the definition of *g*(*q*) is given by:

 (S18)

For the partial derivatives of *g#* (*q*) with *au,k*(*q*) and *xi,k*(*q*), we have:

 (S19)

Then, with the KKT conditions of ∀*mu,k*(*q*), *au,k*(*q*): *mu,k*(*q*)*au,k*(*q*)=0 and ∀*ni,k*(*q*), *xi,k*(*q*): *ni,k*(*q*)*xi,k*(*q*)=0 for (S17), we achieve the following equations based on (S19) [19, 21, 24, 39]:

 (S20)

To satisfy the nonnegativity of output LFs *au,k*(*q*) and *xi,k*(*q*), (S20) can be rewritten as:

 (S21)

Note that (S21) is consistent with the update rules of *au,k*(*q*) and *xi,k*(*q*) based on (4c) and (4d). Therefore, (S17)-(S21) show that learning rules of A2NLF-PSO/A2NLF-TPE are closely connected with the KKT conditions of its learning objective.

Then considering the KKT conditions related to *au,k*(*q*):

 (S22a)

 (S22b)

 (S22c)

 (S22d)

where *a\** *u*,*k*(*q*) is a KKT stationary point of *au,k*(*q*), and *m\** *u*,*k*(*q*) is a limit point of the sequence {*mt* *u*,*k*(*q*)} generated by the update rules of *mu,k* based on (S19). According to (S17)-(S21) and *at* *u*,*k*(*q*)=0, conditions (S22a)-(S22c) are satisfied. Thus, we have:

 (S23)

Thus, we focus on condition (S22d). Since *at* *u*,*k*(*q*)>0 in this case, the update rule for *au,k*(*q*) is given as:

 (S24)

By substituting (S24) into (S23), we have *m\** *u*,*k*(*q*)=0. Hence, conditions (S22c) and (S22d) are fulfilled. Note that as *xt* *i*,*k*(*q*)>0 in this case, the proof regarding the KKT conditions of *xu*,*k*(*q*) can be achieved similarly. *Theorem* 1 stands, and **Step 3** is implemented. □

1. *Proof of Lemma* 3

The difference between  and  in this case is also given by (S5). Considering the fact of *at* *u*,*k*(*q*)=0 and, *xt* *i*,*k*(*q*)>0 the difference between and  is:

 (S25)

Moreover,  and  yields:

 (S26)

where (I) is based on the update rules of (*hu*,*k*(*q*), *wi*,*k*(*q*)) given in (4e), (4f) and (5c), and (II) is achieved with (22b) and *at* *u*,*k*(*q*)=0.

With (S5), (S25) and (S26), we have the following deduction:

 (S27)

Owing to (29), (25a) stands, which indicates that the augmented Lagrangian function (3) related to *s*(*q*) or *τq* is non-increasing as *at* *u*,*k*(*q*)=0 and *xt* *i*,*k*(*q*)>0 in this case. Then after the *t-*th iteration, we substitute *at* *u*,*k*(*q*)=0 into (S10):

 (S28)

(S28) indicates that if (24b) is fulfilled, (25b) holds, thereby making (3) related to *s*(*q*) or *τq* lower-bounded as *at* *u*,*k*(*q*)=0, and *xt* *i*,*k*(*q*)>0 in this case. Based on the above inferences, *Lemma* 2 stands, and **Step 4** is implemented.□

1. *Proof of Theorem* 2

***Part a*.** Following *Lemma* 3,  converges as *t*→∞, indicating that (S11) is fulfilled. With (26), (29) and (S27), we have (S12a), (S12b), (S12d), (S12f) and the following inferences:

 (S29a)

 (S29b)

Then according to (S12f), (S13b) is fulfilled. Hence, based on (S13b), (S29b) and *at* *u*,*k*(*q*)=0, (27) is fulfilled.

***Part b*.** Firstly, considering a limit point  of a sequence  generated by the update rules of {*ψ*1(*q*), *ψ*2(*q*)} based on (4) and (5), according to (27) and (S15), (S16) holds when *at* *u*,*k*(*q*)=0, and *xt* *i*,*k*(*q*)>0 in this case. Then considering the KKT conditions related to *au,k*(*q*), i.e., (S22).

According to (S17)-(S21) and with *at* *u*,*k*(*q*)=0, conditions (S22a)-(S22c) are naturally satisfied. Thus, we focus on analyzing condition (S22d). Since we have *at* *u*,*k*(*q*)=0 in this case, the following inequality holds:

 (S30)

Note that (S30) indicates that:

 (S31)

Thus, condition (S22) are all fulfilled in this case. Note that as *xt* *i*,*k*(*q*)>0 in this case, the proof regarding the KKT conditions of *xu*,*k*(*q*) can be achieved similarly. *Theorem* 2 stands, and **Step 5** is implemented. ∎

# Additional Figures



Fig. S1. Processing flow of an NMF model.



Fig. S2. Processing flow of an NLFA model.

# Additional Tables

TABLE S.I. Optimal Hyper-parameters during M1’s ten times’ training process on D1-4

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Type** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **D1** | **RMSE** | *λ*=2-4, *η*=2 | *λ*=2-4, *η*=2 | *λ*=2-4, *η*=2 | *λ*=2-2, *η*=1 | *λ*=2-4, *η*=2 | *λ*=2-4, *η*=2 | *λ*=2-4, *η*=2 | *λ*=2-4, *η*=2 | *λ*=2-2, *η*=1 | *λ*=2-4, *η*=2 |
| **MAE** | *λ*=2-2, *η*=1 | *λ*=2-4, *η*=1 | *λ*=2-4, *η*=1 | *λ*=2-2, *η*=1 | *λ*=2-4, *η*=1 | *λ*=2-4, *η*=1 | *λ*=2-4, *η*=1 | *λ*=2-2, *η*=1 | *λ*=2-4, *η*=1 | *λ*=2-4, *η*=1 |
| **D2** | **RMSE** | *λ*=1, *η*=1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=1 | *λ*=1, *η*=2 | *λ*=1, *η*=2 | *λ*=1, *η*=1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=1 | *λ*=1, *η*=2 |
| **MAE** | *λ*=1, *η*=2 | *λ*=1, *η*=2 | *λ*=1, *η*=2-1 | *λ*=1, *η*=1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=1 | *λ*=1, *η*=2 | *λ*=1, *η*=2 |
| **D3** | **RMSE** | *λ*=1, *η*=1 | *λ*=1, *η*=2 | *λ*=1, *η*=2-1 | *λ*=1, *η*=2 | *λ*=1, *η*=1 | *λ*=1, *η*=1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=1 | *λ*=1, *η*=1 | *λ*=1, *η*=2 |
| **MAE** | *λ*=1, *η*=1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=2-2 | *λ*=1, *η*=2 | *λ*=1, *η*=2-1 | *λ*=1, *η*=1 | *λ*=1, *η*=2-2 | *λ*=1, *η*=1 | *λ*=1, *η*=2 | *λ*=1, *η*=2 |
| **D4** | **RMSE** | *λ*=1, *η*=2-1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=2 | *λ*=1, *η*=1 | *λ*=1, *η*=2 | *λ*=1, *η*=1 | *λ*=1, *η*=2 | *λ*=1, *η*=1 |
| **MAE** | *λ*=1, *η*=2 | *λ*=1, *η*=1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=1 | *λ*=1, *η*=2 | *λ*=1, *η*=2 | *λ*=1, *η*=1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=2-1 | *λ*=1, *η*=2-1 |

TABLE S.II. Optimal Hyper-parameters during M2’s ten times’ training process on D1-4

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Type** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **D1** | **RMSE** | *λ*=1.6 | *λ*=1.6 | *λ*=1.6 | *λ*=1.8 | *λ*=1.6 | *λ*=1.8 | *λ*=1.6 | *λ*=1.8 | *λ*=1.6 | *λ*=1.6 |
| **MAE** | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.2 | *λ*=1.8 | *λ*=1.8 | *λ*=1.2 |
| **D2** | **RMSE** | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 |
| **MAE** | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 |
| **D3** | **RMSE** | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 |
| **MAE** | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 | *λ*=1.8 |
| **D4** | **RMSE** | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure |
| **MAE** | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure |

1M2 fails to obtain the final results on D4: their memory requirements are too large to meet on our experimental environment as shown in Table VI.

TABLE S.III. Optimal Hyper-parameters during M3’s ten times’ training process on D1-4

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Type** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **D1** | **RMSE** | 𝜇=2-1, *p*=1.9 | 𝜇=2-1, *p*=1.8 | 𝜇=2-1, *p*=1.7 | 𝜇=2-1, *p*=1.8 | 𝜇=2-1, *p*=1.8 | 𝜇=2-1, *p*=1.7 | 𝜇=2-1, *p*=1.8 | 𝜇=2-1, *p*=1.9 | 𝜇=2-1, *p*=1.8 | 𝜇=2-1, *p*=1.8 |
| **MAE** | 𝜇=2-1, *p*=1.8 | 𝜇=2-1, *p*=1.9 | 𝜇=2-1, *p*=1.8 | 𝜇=2-1, *p*=1.7 | 𝜇=2-1, *p*=1.8 | 𝜇=2-1, *p*=1.7 | 𝜇=2-1, *p*=1.8 | 𝜇=2-1, *p*=1.8 | 𝜇=2-1, *p*=1.9 | 𝜇=2-1, *p*=1.9 |
| **D2** | **RMSE** | 𝜇=2, *p*=1.9 | 𝜇=4, *p*=1.9 | 𝜇=4 *p*=1.9 | 𝜇=2-1, *p*=1.8 | 𝜇=2-1, *p*=1.9 | 𝜇=2, *p*=1.9 | 𝜇=2-1, *p*=1.8 | 𝜇=2, *p*=1.9 | 𝜇=2, *p*=1.8 | 𝜇=2, *p*=1.8 |
| **MAE** | 𝜇=4, *p*=1.9 | 𝜇=2, *p*=1.9 | 𝜇=2, *p*=1.8 | 𝜇=4, *p*=1.9 | 𝜇=4, *p*=1.9 | 𝜇=2, *p*=1.9 | 𝜇=1, *p*=1.9 | 𝜇=2, *p*=1.8 | 𝜇=2, *p*=1.8 | 𝜇=2, *p*=1.9 |
| **D3** | **RMSE** | 𝜇=4, *p*=1.9 | 𝜇=4, *p*=1.9 | 𝜇=2, *p*=1.8 | 𝜇=2, *p*=1.8 | 𝜇=4, *p*=1.9 | 𝜇=4, *p*=1.9 | 𝜇=2, *p*=1.9 | 𝜇=2, *p*=1.9 | 𝜇=2, *p*=1.9 | 𝜇=2, *p*=1.9 |
| **MAE** | 𝜇=2, *p*=1.9 | 𝜇=2, *p*=1.8 | 𝜇=2, *p*=1.9 | 𝜇=2, *p*=1.8 | 𝜇=4, *p*=1.9 | 𝜇=4, *p*=1.9 | 𝜇=4, *p*=1.9 | 𝜇=2, *p*=1.8 | 𝜇=2, *p*=1.9 | 𝜇=2, *p*=1.9 |
| **D4** | **RMSE** | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure |
| **MAE** | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure |

1M3 fails to obtain the final results on D4: their memory requirements are too large to meet on our experimental environment as shown in Table VI.

TABLE S.IV. Optimal Hyper-parameters during M4’s ten times’ training process on D1-4

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Type** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **D1** | **RMSE** | *α*=104, *β*=103 | *α*=104, *β*=103 | *α*=104, *β*=103 | *α*=103, *β*=104 | *α*=104, *β*=103 | *α*=103, *β*=104 | *α*=104, *β*=103 | *α*=104, *β*=103 | *α*=104, *β*=103 | *α*=104, *β*=103 |
| **MAE** | *α*=104, *β*=103 | *α*=103, *β*=104 | *α*=104, *β*=103 | *α*=104, *β*=103 | *α*=104, *β*=103 | *α*=104, *β*=103 | *α*=104, *β*=103 | *α*=104, *β*=103 | *α*=103, *β*=104 | *α*=103, *β*=104 |
| **D2** | **RMSE** | *α*=103, *β*=104 | *α*=104, *β*=102 | *α*=103, *β*=103 | *α*=103, *β*=104 | *α*=103, *β*=103 | *α*=104, *β*=102 | *α*=103, *β*=103 | *α*=103, *β*=104 | *α*=104, *β*=102 | *α*=103, *β*=103 |
| **MAE** | *α*=103, *β*=104 | *α*=103, *β*=104 | *α*=103, *β*=103 | *α*=103, *β*=104 | *α*=103, *β*=103 | *α*=104, *β*=103 | *α*=103, *β*=103 | *α*=103, *β*=104 | *α*=103, *β*=104 | *α*=103, *β*=103 |
| **D3** | **RMSE** | *α*=104, *β*=103 | *α*=103, *β*=104 | *α*=104, *β*=103 | *α*=103, *β*=104 | *α*=103, *β*=104 | *α*=104, *β*=103 | *α*=104, *β*=103 | *α*=103, *β*=104 | *α*=103, *β*=104 | *α*=104, *β*=103 |
| **MAE** | *α*=103, *β*=104 | *α*=103, *β*=104 | *α*=104, *β*=103 | *α*=104, *β*=103 | *α*=104, *β*=103 | *α*=103, *β*=104 | *α*=103, *β*=104 | *α*=103, *β*=104 | *α*=104, *β*=103 | *α*=103, *β*=104 |
| **D4** | **RMSE** | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure |
| **MAE** | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure |

1M4 fails to obtain the final results on D4: their memory requirements are too large to meet on our experimental environment as shown in Table VI.

TABLE S.V. Optimal Hyper-parameters during M5’s ten times’ training process on D1-4

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Type** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **D1** | **RMSE** | *ρ*=2-5 | *ρ*=2-5 | *ρ*=2-5 | *ρ*=2-5 | *ρ*=2-5 | *ρ*=2-5 | *ρ*=2-5 | *ρ*=2-5 | *ρ*=2-5 | *ρ*=2-5 |
| **MAE** | *ρ*=2-7 | *ρ*=2-5 | *ρ*=2-7 | *ρ*=2-5 | *ρ*=2-5 | *ρ*=2-7 | *ρ*=2-5 | *ρ*=2-5 | *ρ*=2-7 | *ρ*=2-5 |
| **D2** | **RMSE** | *ρ*=2-15 | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-15 | *ρ*=2-15 | *ρ*=2-13 |
| **MAE** | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-15 | *ρ*=2-15 | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-13 | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-15 |
| **D3** | **RMSE** | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-13 | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-13 |
| **MAE** | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-13 | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-13 | *ρ*=2-13 | *ρ*=2-15 | *ρ*=2-15 | *ρ*=2-13 |
| **D4** | **RMSE** | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure |
| **MAE** | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure | 1Failure |

1M5 fails to obtain the final results on D4: their memory requirements are too large to meet on our experimental environment as shown in Table VI.

TABLE S.VI. Optimal Hyper-parameters during M6’s ten times’ training process on D1-4

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Type** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **D1** | **RMSE** | *λ*=10-4, *η*=5×10-3, 1bs=512 | *λ*=10-3, *η*=5×10-3, bs=512 | *λ*=10-5, *η*=10-3, bs=512 | *λ*=10-3, *η*=5×10-3, bs=512 | *λ*=10-4, *η*=5×10-3, bs=512 | *λ*=10-5, *η*=10-3, bs=512 | *λ*=10-3, *η*=5×10-3, bs=512 | *λ*=10-5, *η*=10-3, bs=512 | *λ*=10-4, *η*=5×10-3, bs=512 | *λ*=10-3, *η*=5×10-3, bs=512 |
| **MAE** | *λ*=10-5, *η*=10-3, bs=512 | *λ*=10-5, *η*=10-3, bs=512 | *λ*=10-2, *η*=5×10-3, bs=512 | *λ*=10-2, *η*=5×10-3, bs=512 | *λ*=10-2, *η*=5×10-3, bs=512 | *λ*=10-5, *η*=10-3, bs=512 | *λ*=10-5, *η*=10-3, bs=512 | *λ*=10-5, *η*=10-3, bs=512 | *λ*=10-5, *η*=10-3, bs=512 | *λ*=10-2, *η*=5×10-3, bs=512 |
| **D2** | **RMSE** | *λ*=10-4, *η*=10-3, bs=64 | *λ*=10-4, *η*=10-3, bs=64 | *λ*=10-1, *η*=5×10-3, bs=64 | *λ*=10-1, *η*=5×10-3, bs=64 | *λ*=10-2, *η*=10-3, bs=64 | *λ*=10-1, *η*=5×10-3, bs=64 | *λ*=10-2, *η*=10-3, bs=64 | *λ*=10-2, *η*=10-3, bs=64 | *λ*=10-1, *η*=5×10-3, bs=64 | *λ*=10-4, *η*=10-3, bs=64 |
| **MAE** | *λ*=10-4, *η*=10-2, bs=64 | *λ*=10-3, *η*=10-2, bs=64 | *λ*=10-4, *η*=10-2, bs=64 | *λ*=10-3, *η*=10-2, bs=64 | *λ*=10-4, *η*=10-2, bs=64 | *λ*=10-2, *η*=10-3, bs=64 | *λ*=10-4, *η*=10-2, bs=64 | *λ*=10-2, *η*=10-3, bs=64 | *λ*=10-2, *η*=10-3, bs=64 | *λ*=10-2, *η*=10-3, bs=64 |
| **D3** | **RMSE** | *λ*=10-4, *η*=5×10-4, bs=128 | *λ*=10-4, *η*=5×10-4, bs=128 | *λ*=10-4, *η*=5×10-4, bs=128 | *λ*=10-4, *η*=5×10-4, bs=128 | *λ*=10-4, *η*=5×10-4, bs=128 | *λ*=10-5, *η*=10-3, bs=128 | *λ*=10-5, *η*=10-3, bs=128 | *λ*=10-4, *η*=5×10-4, bs=128 | *λ*=10-4, *η*=5×10-4, bs=128 | *λ*=10-4, *η*=5×10-4, bs=128 |
| **MAE** | *λ*=10-5, *η*=10-3, bs=128 | *λ*=10-4, *η*=10-3, bs=128 | *λ*=10-5, *η*=10-3, bs=128 | *λ*=10-5, *η*=10-3, bs=128 | *λ*=10-4, *η*=10-3, bs=128 | *λ*=10-4, *η*=10-3, bs=128 | *λ*=10-4, *η*=10-3, bs=128 | *λ*=10-4, *η*=10-3, bs=128 | *λ*=10-4, *η*=10-3, bs=128 | *λ*=10-4, *η*=10-3, bs=128 |
| **D4** | **RMSE** | *λ*=10-3, *η*=10-4, bs=512 | *λ*=10-4, *η*=10-3, bs=512 | *λ*=10-5, *η*=5×10-3, bs=512 | *λ*=10-4, *η*=10-3, bs=512 | *λ*=10-4, *η*=10-3, bs=512 | *λ*=10-3, *η*=10-4, bs=512 | *λ*=10-5, *η*=5×10-3, bs=512 | *λ*=10-3, *η*=10-4, bs=512 | *λ*=10-5, *η*=5×10-3, bs=512 | *λ*=10-3, *η*=10-4, bs=512 |
| **MAE** | *λ*=10-2, *η*=10-3, bs=512 | *λ*=10-2, *η*=10-3, bs=512 | *λ*=10-2, *η*=5×10-3, bs=512 | *λ*=10-2, *η*=5×10-3, bs=512 | *λ*=10-5, *η*=5×10-3, bs=512 | *λ*=10-2, *η*=5×10-3, bs=512 | *λ*=10-5, *η*=5×10-3, bs=512 | *λ*=10-2, *η*=5×10-3, bs=512 | *λ*=10-2, *η*=10-3, bs=512 | *λ*=10-2, *η*=5×10-3, bs=512 |

1The abbreviation ‘bs’ denotes the batch size adopted by M6 on an HDI matrix.

TABLE S.VII. Optimal Hyper-parameters during M7’s ten times’ training process on D1-4

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Type** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **D1** | **RMSE** | *λ*=10-3, *η*=10-3, 1bs=2048 | *λ*=10-3, *η*=5×10-3, bs=2048 | *λ*=10-3, *η*=5×10-3, bs=2048 | *λ*=10-3, *η*=5×10-2, bs=2048 | *λ*=10-3, *η*=5×10-3, bs=2048 | *λ*=5×10-3, *η*=5×10-2, bs=2048 | *λ*=10-3, *η*=5×10-3, bs=2048 | *λ*=5×10-3, *η*=5×10-2, bs=2048 | *λ*=10-3, *η*=5×10-2, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 |
| **MAE** | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=5×10-3, *η*=10-3, bs=2048 | *λ*=10-3 *η*=5×10-3, bs=2048 | *λ*=10-3, *η*=5×10-3, bs=2048 | *λ*=5×10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=5×10-2, bs=2048 |
| **D2** | **RMSE** | *λ*=10-3, *η*=5×10-3, bs=1024 | *λ*=10-3, *η*=10-3, bs=1024 | *λ*=10-3, *η*=10-3, bs=1024 | *λ*=10-3, *η*=5×10-3, bs=1024 | *λ*=10-3, *η*=5×10-3, bs=1024 | *λ*=10-3, *η*=10-3, bs=1024 | *λ*=10-3, *η*=5×10-3, bs=1024 | *λ*=10-3, *η*=10-3, bs=1024 | *λ*=10-3, *η*=5×10-3, bs=1024 | *λ*=10-3, *η*=10-3, bs=1024 |
| **MAE** | *λ*=10-3, *η*=5×10-3, bs=1024 | *λ*=10-3, *η*=10-3, bs=1024 | *λ*=10-3, *η*=10-3, bs=1024 | *λ*=10-3, *η*=5×10-3, bs=1024 | *λ*=10-3, *η*=10-3, bs=1024 | *λ*=10-3, *η*=5×10-3, bs=1024 | *λ*=10-3, *η*=10-3, bs=1024 | *λ*=5×10-3, *η*=10-3, bs=1024 | *λ*=10-3, *η*=5×10-3, bs=1024 | *λ*=10-3, *η*=10-3, bs=1024 |
| **D3** | **RMSE** | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=5×10-3, bs=2048 | *λ*=5×10-3, *η*=5×10-2, bs=2048 | *λ*=5×10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=5×10-3, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=5×10-2, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 |
| **MAE** | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=5×10-3, bs=2048 | *λ*=10-3, *η*=5×10-2, bs=2048 | *λ*=10-3, *η*=5×10-3, bs=2048 | *λ*=10-3, *η*=5×10-3, bs=2048 | *λ*=5×10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 | *λ*=10-3, *η*=10-3, bs=2048 |
| **D4** | **RMSE** | *λ*=10-3, *η*=10-3, bs=4096 | *λ*=10-3, *η*=5×10-3, bs=4096 | *λ*=10-3, *η*=5×10-3, bs=4096 | *λ*=10-3, *η*=10-3, bs=4096 | *λ*=10-3, *η*=5×10-2, bs=4096 | *λ*=10-3, *η*=10-3, bs=4096 | *λ*=10-3, *η*=5×10-3, bs=4096 | *λ*=10-3, *η*=10-3, bs=4096 | *λ*=10-3, *η*=5×10-3, bs=4096 | *λ*=10-3, *η*=10-3, bs=4096 |
| **MAE** | *λ*=10-3, *η*=10-3, bs=4096 | *λ*=10-3, *η*=10-3, bs=4096 | *λ*=10-3, *η*=5×10-3, bs=4096 | *λ*=5×10-3, *η*=10-3, bs=4096 | *λ*=10-3, *η*=10-3, bs=4096 | *λ*=10-3, *η*=10-3, bs=4096 | *λ*=5×10-3, *η*=5×10-3, bs=4096 | *λ*=10-3, *η*=10-3, bs=4096 | *λ*=10-3, *η*=5×10-3, bs=4096 | *λ*=10-3, *η*=10-3, bs=4096 |

1The abbreviation ‘bs’ denotes batch size adopted by M7 on an HDI matrix.

TABLE S.VIII. RMSE, MAE and Time Cost of M1, M8 and M9 on D1-4.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Model** | **Case** | **Prediction Accuracy** | 1**Tuning Time Cost (Secs)** | 2**Testing Time Cost (Secs)** | **Total Time Cost (Secs)** |
| D1 | M1 | RMSE | 0.2373±2.2E-6 | 428±22.7 | 6±2.4 | 434±25 |
| MAE | 0.1815±1.1E-6 | 439±25.4 | 6±2.8 | 445±28 |
| M8 | RMSE | **0.2339±2.7E-4** | - | - | **46±4** |
| MAE | **0.1793±3.3E-4** | - | - | **60±5** |
| M9 | RMSE | 0.2573±1.4E-6 | - | - | 387±33 |
| MAE | 0.2015±2.8E-6 | - | - | 396±46 |
| D2 | M1 | RMSE | 1.0187±1.1E-6 | 271±45.4 | 4±1.7 | 275±47 |
| MAE | 0.8079±2.5E-6 | 305±38.3 | 3±0.9 | 308±39 |
| M8 | RMSE | **1.0158±6.4E-4** | - | - | **25±5** |
| MAE | **0.7848±9.3E-4** | - | - | **29±8** |
| M9 | RMSE | 1.0209±2.0E-5 | - | - | 87±11 |
| MAE | 0.7883±1.3E-5 | - | - | 85±9 |
| D3 | M1 | RMSE | **0.8665±7.8E-4** | 739±72.3 | 21±14.7 | 759±87 |
| MAE | 0.6829±1.7E-6 | 756±53.8 | 3±0.5 | 763±54 |
| M8 | RMSE | 0.8673±1.3E-3 | - | - | **26±6** |
| MAE | **0.6785±9.6E-5** | - | - | **30±9** |
| M9 | RMSE | 0.8684±4.7E-4 | - | - | 487±59 |
| MAE | 0.6793±2.8E-5 | - | - | 268±40 |
| D4 | M1 | RMSE | 0.8096±2.9E-6 | 2,934±353.8 | 38±24.6 | 3,972±378 |
| MAE | 0.6221±7.9E-7 | 9,203±828.9 | 33±20.4 | 9,236±849 |
| M8 | RMSE | 0.8091±3.0E-3 | - | - | **334±46** |
| MAE | **0.6191±1.7E-4** | - | - | **358±31** |
| M9 | RMSE | **0.8086±6.0E-5** | - | - | 3731±311 |
| MAE | 0.6199±7.1E-6 | - | - | 5532±566 |

1Time cost consumed by M1 for manually grid-searching optimal hyper-parameters; 2Time cost consumed by M1 with obtained hyper-parameters.

TABLE S.IX. RMSE/MAE of M1-9 on D1-4, including Win/Loss counts and Friedman Rank, where ⚫ indicates that both M8 and M9 have higher RMSE/MAE than the rival models

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Case** | **M1** | **M2** | **M3** | **M4** | **M5** | **M6** | **M7** | **M8** | **M9** |
| D1 | RMSE | 0.2373±2.2E-6 | 0.3058±5.0E-4 | 0.3047±4.3E-5 | 0.2384±1.0E-4 | 0.4913±1.0E-4 | ⚫**0.2302±2.6E-3** | 0.2504±4.7E-4 | 0.2339±2.7E-4 | 0.2573±1.4E-6 |
| MAE | 0.1815±1.1E-6 | 0.2439±4.4E-5 | 0.2422±3.2E-5 | 0.1832±4.8E-5 | 0.4111±1.6E-2 | ⚫**0.1792±7.9E-5** | 0.1953±8.9E-4 | 0.1793±3.3E-4 | 0.2015±2.8E-6 |
| D2 | RMSE | 1.0187±1.1E-6 | 1.1281±7.2E-3 | 1.1257±1.8E-4 | 1.0787±8.6E-7 | 1.8808±2.8E-2 | 1.1494±4.1E-5 | 1.0255±4.5E-4 | **1.0158±6.4E-4** | 1.0209±2.0E-5 |
| MAE | 0.8079±2.5E-6 | 0.9256±1.5E-3 | 0.9229±1.9E-4 | 0.8580±6.2E-6 | 1.5403±2.6E-2 | 0.9257±2.1E-4 | 0.7945±1.6E-3 | **0.7848±9.3E-4** | 0.7883±1.3E-5 |
| D3 | RMSE | ⚫0.8665±7.8E-4 | 1.0336±5.4E-4 | 1.0848±5.3E-5 | 0.8713±3.7E-4 | 2.2963±9.8E-3 | 0.8845±1.1E-3 | 0.8972±8.2E-3 | 0.8673±1.3E-3 | 0.8684±4.7E-4 |
| MAE | 0.6829±1.7E-6 | 0.8832±4.0E-4 | 0.9021±4.2E-5 | 0.6802±3.3E-4 | 1.9190±9.6E-3 | 0.7021±5.5E-3 | 0.7036±3.6E-3 | **0.6785±9.6E-5** | 0.6793±2.8E-5 |
| D4 | RMSE | 0.8096±2.9E-6 | 1Failure | 1Failure | 1Failure | 1Failure | 0.8436±1.2E-3 | 0.8574±1.2E-3 | 0.8091±3.0E-3 | **0.8086±6.0E-5** |
| MAE | 0.6221±7.9E-7 | 1Failure | 1Failure | 1Failure | 1Failure | 0.6530±4.9E-3 | 0.6647±3.3E-3 | **0.6191±1.7E-4** | 0.6199±7.1E-6 |
| **Win/Loss** | | 8/1 | 9/0 | 9/0 | 9/0 | 9/0 | 7/2 | 9/0 | - | - |
| **F-Rank** | | 2.875 | 7.375 | 7.125 | 5 | 8.625 | 4.5 | 4.875 | **1.5** | 3.125 |

1M2-M5 fails to obtain the final results on D4 on our experimental environment as shown in Table VI.

TABLE S.X. Total time cost of M1-9 in RMSE/MAE on D1-4 (Secs), including Win/Loss counts and Friedman Rank

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Case** | **M1** | **M2** | **M3** | **M4** | **M5** | **M6** | **M7** | **M8** | **M9** |
| D1 | RMSE | 434±25 | 38,499±2,332 | 43,271±2,962 | 209,670±36,625 | 3,856±578 | 10,804±832 | 2,011,581±481,806 | **46±4** | 387±33 |
| MAE | 445±28 | 51,443±2,643 | 189,090±29,135 | 210,745±27,190 | 3,912±103 | 13,432±2,219 | 1,809,472±393,344 | **60±5** | 396±46 |
| D2 | RMSE | 275±47 | 236±35 | 128±21 | 138±39 | 55±7 | 2,109±295 | 888,704±223,355 | **25±5** | 87±11 |
| MAE | 308±39 | 32±3 | 128±27 | 766±64 | 56±6 | 2,256±633 | 2,015,970±861,492 | **29±8** | 85±9 |
| D3 | RMSE | 759±87 | 42,578±3,258 | 71,927±2,718 | 45,849±2,479 | 717±22 | 8,359±532 | 883,008±63,588 | **26±6** | 487±59 |
| MAE | 763±54 | 1,906±291 | 12,294±2,478 | 48,821±2,363 | 716±53 | 7,700±219 | 676,921±41,748 | **30±9** | 268±40 |
| D4 | RMSE | 3,972±378 | 1Failure | 1Failure | 1Failure | 1Failure | 266,885±12,775 | 45,533,828±8,105,384 | **334±46** | 3731±311 |
| MAE | 9,236±849 | 1Failure | 1Failure | 1Failure | 1Failure | 232,974±13,157 | 50,643,155±6,301,613 | **358±31** | 5532±566 |
| **Win/Loss** | | 9/0 | 9/0 | 9/0 | 9/0 | 9/0 | 9/0 | 9/0 | - | - |
| **F-Rank** | | 4.125 | 5.75 | 6.625 | 7.25 | 4.25 | 5.625 | 8 | **1** | 2.375 |

1M2-M5 fails to obtain the final results on D4 on our experimental environment as shown in Table VI.