We perform the **grid search** on  $\{8, 16, 32, 64, 128\}$  for the **latent dimension** and  $\{0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001\}$  for the **learning rate**, and  $\{1e-5, 1e-6, 1e-7, 1e-8, 1e-9, 1e-10, 1e-11, 1e-12, 1e-13, 1e-14, 1e-15\}$  for the **regularization term**. In CJMF,  $\alpha$ ,  $\beta$  and C are fixed as 220000, 0.5 and 8, respectively. As a result, the best hyperparameters are chosen as the following table. Like before, we apply early-stopping on the validation set for determining the best epoch.

					Models			
Datasets	Hyperparameters	MF	Rel-MF	MF-DU	AT	AT-DU	CJMF	CJMF-DU
ML100K	hidden dim.	64	64	64	64	64	64	64
	weight decay	1.00E-06	1.00E-06	1.00E-08	1.00E-10	1.00E-08	1.00E-09	1.00E-11
	learning rate	0.001	0.001	0.001	0.0001	0.0001	0.001	0.001
	batch size	1024	1024	1024	1024	1024	1024	1024
Coat	hidden dim.	128	128	128	128	128	128	128
	weight decay	1.00E-09	1.00E-05	1.00E-07	1.00E-10	1.00E-11	1.00E-09	1.00E-11
	learning rate	0.005	0.005	0.01	0.0005	0.001	0.005	0.01
	batch size	1024	1024	1024	1024	1024	1024	1024
Yahoo! R3	hidden dim.	64	64	64	64	64	64	64
	weight decay	1.00E-07	1.00E-07	1.00E-08	1.00E-14	1.00E-12	1.00E-07	1.00E-09
	learning rate	0.001	0.001	0.001	0.0001	0.0001	0.001	0.005
	batch size	1024	1024	1024	1024	1024	1024	1024