3.a
No.

3.b
No.

3.c
In the shared model setting,
the model trained with λ_F = 0.99 and λ_R = 0.01.
outperforms
the model trained with λ_F = 0.5 and λ_R = 0.5.

I think the reason is:

1. When the model parameters (U and Q) are shared across 2 tasks, the gradients caused by one task will be applied on another task. Therefore, both 2 tasks tie together much more closely than the case when there is no sharing.
2. When both 2 tasks tie closely, we should give more weight to the task with higher importance. In this case, the factorization task has higher importance than the regression task. If the model does well on the factorization task, it means that the model can accurately predict which movie a user would like to watch. Intuitively, such a model is supposed to do well to do the score assignment task, as well.

Therefore, the model trained with more weight on factorization tasks should perform better, in the shared model setting.