This is some implemetation note

The grouping of output features

Six head, QA tangled:

[Group 1]:

- question_multi_intent
- question_type_choice
- question_type_reason_explanation
- answer_type_reason_explanation

[Group 2]:

- question_asker_intent_understanding
- question_body_critical
- question_interestingness_others
- question_interestingness_self
- question_well_written

[Group 3]:

- question_conversational
- question_not_really_a_question
- question_opinion_seeking
- question_type_compare
- question_type_consequence
- question_type_definition
- question_type_entity
- question_type_spelling

[Group 4]:

- answer_helpful
- answer_level_of_information
- answer_plausible
- answer_relevance
- answer_satisfaction
- answer_well_written

[Group 5]:

- question_fact_seeking
- question_type_procedure
- answer_type_procedure

[Group 6]:

- question_expect_short_answer
- question_has_commonly_accepted_answer
- question_type_instructions
- answer_type_instructions

Six head, QA splited:

[Group 1]:

- question_expect_short_answer
- question_fact_seeking
- question_has_commonly_accepted_answer
- question_type_instructions
- question_type_procedure

[Group 2]:

- question_asker_intent_understanding
- question_body_critical
- question_interestingness_others
- question_interestingness_self
- question_well_written

[Group 3]:

- question_conversational
- question_opinion_seeking

[Group 4]:

- question_multi_intent
- question_not_really_a_question
- question_type_choice
- question_type_compare
- question_type_consequence
- question_type_definition
- question_type_entity
- question_type_reason_explanation
- question_type_spelling

[Group 1]:

- answer_type_instructions
- answer_type_procedure

[Group 2]:

- answer_helpful
- answer_level_of_information
- answer_plausible
- answer_relevance
- answer_satisfaction
- answer_type_reason_explanation
- answer_well_written

The performance of different projector head grouping strategy - With 6 head QA distangled later

Below experiment is done under the setting of one Deberta and Kflod

Stragegy	Fold0Epoch2(Train)	Fold1Epoch2(Train)	Fold2Epoch2(Train)	Fold3Epoch2(Train)	Fold4Epoch2(Train)	Public(Eval)	Private(Eval)
One global head						0.30509	0.28049
Two head, QA splited	Loss: 0.3770 - Raw Score: 0.3391	Loss: 0.3797 - Raw Score: 0.3287	Loss: 0.3779 - Raw Score: 0.3360	Loss: 0.3793 - Raw Score: 0.3207	Loss: 0.3788 - Raw Score: 0.3283	0.31465	0.29607
Six head, global	Loss: 0.3750 - Raw Score: 0.3413	Loss: 0.3775 - Raw Score: 0.3363	Loss: 0.3767 - Raw Score: 0.3392	Loss: 0.3783 - Raw Score: 0.3239	Loss: 0.3733 - Raw Score: 0.3449	0.31958	0.30399
Six head, QA splited	Loss: 0.3730 - Raw Score: 0.3525	Loss: 0.3776 - Raw Score: 0.3354	Loss: 0.3767 - Raw Score: 0.3333	Loss: 0.3788 - Raw Score: 0.3285	Loss: 0.3784 - Raw Score: 0.3295	0.31968	0.30270

The use of class weight CE loss - With class weight later

Below experiment is done under the setting of one Deberta and six head, QA splited

Stragegy	Fold0Epoch2(Train)	Fold1Epoch2(Train)	Fold2Epoch2(Train)	Fold3Epoch2(Train)	Fold4Epoch2(Train)	Public(Eval)	Private(Eval)
Without weight	Loss: 0.3730 - Raw Score: 0.3525	Loss: 0.3776 - Raw Score: 0.3354	Loss: 0.3767 - Raw Score: 0.3333	Loss: 0.3788 - Raw Score: 0.3285	Loss: 0.3784 - Raw Score: 0.3295	0.31968	0.30270
With weight	Loss: 0.3964 - Raw Score: 0.3555	Loss: 0.4015 - Raw Score: 0.3331	Loss: 0.3990 - Raw Score: 0.3412	Epoch 2 - Loss: 0.4013 - Raw Score: 0.3379	Epoch 2 - Loss: 0.3999 - Raw Score: 0.3431	0.32706	0.31128

The use of larger lr for head - With larger lr for head later

Stragegy	Fold0Epoch2(Train)	Fold1Epoch2(Train)	Fold2Epoch2(Train)	Fold3Epoch2(Train)	Fold4Epoch2(Train)	Public(Eval)	Private(Eval)
Smaller LR	Loss: 0.3730 - Raw Score: 0.3525	Loss: 0.3776 - Raw Score: 0.3354	Loss: 0.3767 - Raw Score: 0.3333	Loss: 0.3788 - Raw Score: 0.3285	Loss: 0.3784 - Raw Score: 0.3295	0.31968	0.30270

Stragegy	Fold0Epoch2(Train)	Fold1Epoch2(Train)	Fold2Epoch2(Train)	Fold3Epoch2(Train)	Fold4Epoch2(Train)	Public(Eval)	Private(Eval)
Larger LR	Loss: 0.3894 - Raw Score: 0.3876	Loss: 0.3905 - Raw Score: 0.3744	Loss: 0.3913 - Raw Score: 0.3759	Loss: 0.3911 - Raw Score: 0.3653	Epoch 2 - Loss: 0.3906 - Raw Score: 0.3718	0.36439	0.34443

The use of larger epoch - With 5 epochs later

Stragegy	Public(Eval)	Private(Eval)
2 epoch	0.36439	0.34443
5 epoch	0.39181	0.37820

The use of group k fold - Without group k fold when doing experiment

Stragegy	Public(Eval)	Private(Eval)
5 fold	0.39181	0.37820
no fold	0.38451	0.37011

The strategy of head output - Using single regression

https://manikanthgoud123.medium.com/google-quest-q-a-labeling-kaggle-competition-d205bea1e026 question_asker_inten: no. of unique label values: 9 question_body_critic: no. of unique label values: 9 question_conversatio: no. of unique label values: 5 question_expect_shor: no. of unique label values: 5 question_fact_seekin: no. of unique label values: 5 question_has_commonl: no. of unique label values: 5 question_interesting: no. of unique label values: 9 question_interesting: no. of unique label values: 9 question_multi_inten: no. of unique label values: 5 question_not_really_: no. of unique label values: 5 question_opinion_see: no. of unique label values: 5 question_type_choice: no. of unique label values: 5 question_type_compar: no. of unique label values: 5 question_type_conseq: no. of unique label values: 5 question_type_defini: no. of unique label values: 5 question_type_entity: no. of unique label values: 5 question_type_instru: no. of unique label values: 5 question_type_proced: no. of unique label values: 5 question_type_reason: no. of unique label values: 5 question_type_spelli: no. of unique label values: 3 question_well_writte: no. of unique label values: 9 answer_helpful: no. of unique label values: 9 answer_level_of_info: no. of unique label values: 9 answer_plausible: no. of unique label values: 9 answer_relevance: no. of unique label values: 9 answer_satisfaction: no. of unique label values: 17 answer_type_instruct: no. of unique label values: 5 answer_type_procedur: no. of unique label values: 5 answer_type_reason_e: no. of unique label values: 5 answer_well_written: no. of unique label values: 9

No class weight for ordinal regression

Stragegy	Epoch1	Epoch2	Epoch3	Epoch4	Epoch5	Public(Eval)	Private(Eval)
Single regression	Loss: 0.4444 - Raw Score: 0.3487	Loss: 0.3919 - Raw Score: 0.3874	Loss: 0.3777 - Raw Score: 0.3994	Loss: 0.3674 - Raw Score: 0.4040	Loss: 0.3594 - Raw Score: 0.4040	0.38451	0.37011
2rd placed Ordinal regression	Loss: 0.3826 - Spearman Score: 0.2922	Loss: 0.3043 - Spearman Score: 0.3473	Loss: 0.2894 - Spearman Score: 0.3700	Loss: 0.2773 - Spearman Score: 0.3843	Loss: 0.2680 - Spearman Score: 0.3875	0.36534	0.33958

Now we observe the performance of the model on each output feature

Target Column	Spearman of Ordinal	Spearman of single regression
question_asker_intent_understanding	0.4379	0.4362
question_body_critical	0.7300	0.6705
question_conversational	0.4502	0.4656
question_expect_short_answer	0.3493	0.4087
question_fact_seeking	0.4553	0.4856
question_has_commonly_accepted_answer	0.4977	0.5335
question_interestingness_others	0.4191	0.4112
question_interestingness_self	0.5719	0.5479
question_multi_intent	0.5726	0.6488
question_not_really_a_question	0.0642	0.1240
question_opinion_seeking	0.5786	0.6103
question_type_choice	0.7606	0.7873
question_type_compare	0.3049	0.3971
question_type_consequence	0.1575	0.2126

Target Column	Spearman of Ordinal	Spearman of single regression
question_type_definition	0.3422	0.3788
question_type_entity	0.4238	0.5027
question_type_instructions	0.8182	0.8272
question_type_procedure	0.3810	0.4687
question_type_reason_explanation	0.7151	0.7519
question_type_spelling	0.0587	0.0677
question_well_written	0.5785	0.5832
answer_helpful	0.2733	0.3074
answer_level_of_information	0.3714	0.4479
answer_plausible	0.1997	0.2132
answer_relevance	0.2279	0.2286
answer_satisfaction	0.3463	0.3812
answer_type_instructions	0.7983	0.8111
answer_type_procedure	0.3207	0.4337
answer_type_reason_explanation	0.7261	0.7728
answer_well_written	0.2180	0.2722
AVERAGE	0.4383	0.4729

The post processing rounding trick - 4th placed Voters

Stragegy	Public(Eval)	Private(Eval)
No processing	0.36771	0.35538
3rd placed OptimizedRounder	0.38451	0.37011
4th placed Voters	0.39557	0.37100
1st placed Distribution	0.38678	0.36623

Target Column	Raw	OptRound	Voters	Dist.
question_asker_intent_understanding	0.4362	0.4362	0.4011	0.4033
question_body_critical	0.6705	0.6705	0.6654	0.6648
question_conversational	0.4656	0.6079	0.6313	0.6278
question_expect_short_answer	0.4087	0.4087	0.3763	0.3687
question_fact_seeking	0.4856	0.4856	0.4623	0.4593
question_has_commonly_accepted_answer	0.5335	0.5828	0.5412	0.5601
question_interestingness_others	0.4112	0.4112	0.3812	0.3952
question_interestingness_self	0.5479	0.5479	0.5377	0.5437
question_multi_intent	0.6488	0.6496	0.6424	0.6480
question_not_really_a_question	0.1240	0.2289	0.2313	0.2355
question_opinion_seeking	0.6103	0.6103	0.5911	0.5838
question_type_choice	0.7873	0.8111	0.8017	0.7987
question_type_compare	0.3971	0.6808	0.6717	0.6727
question_type_consequence	0.2126	0.4183	0.4214	0.4171
question_type_definition	0.3788	0.7736	0.7495	0.7652
question_type_entity	0.5027	0.7308	0.7189	0.7180
question_type_instructions	0.8272	0.8363	0.8295	0.8277
question_type_procedure	0.4687	0.4689	0.4362	0.3945
question_type_reason_explanation	0.7519	0.7520	0.7384	0.7411

Target Column	Raw	OptRound	Voters	Dist.
question_type_spelling	0.0677	0.3476	0.0677	0.1806
question_well_written	0.5832	0.5832	0.5616	0.5741
answer_helpful	0.3074	0.3074	0.2681	0.2974
answer_level_of_information	0.4479	0.4479	0.3959	0.4303
answer_plausible	0.2132	0.2133	0.1787	0.1949
answer_relevance	0.2286	0.2456	0.2346	0.2248
answer_satisfaction	0.3812	0.3812	0.3728	0.3734
answer_type_instructions	0.8111	0.8130	0.8061	0.8071
answer_type_procedure	0.4337	0.4340	0.3955	0.3644
answer_type_reason_explanation	0.7728	0.7728	0.7585	0.7562
answer_well_written	0.2722	0.2722	0.2048	0.2602
AVERAGE	0.4729	0.5310	0.5024	0.5096

The use of differnet models

No k fold, 5 epochs

Model	Public(Eval)	Private(Eval)
deberta-v3-base		
Llama		
Qwen		
roberta		
Mistral		
Phi		
bge		
instructor		
E5		

The combinations of final models

5 folds, 10 epochs

Model	Public(Eval)	Private(Eval)
deberta-v3-base		