valid (top left): Precision vs Recall
- 0: AUC=0.88 F1=0.82
- 1: AUC=0.86 F1=0.8
- 2: AUC=0.74 F1=0.65
- 3: AUC=0.66 F1=0.64
- 4: AUC=0.68 F1=0.64

test (top right): Precision vs Recall
- 0: AUC=0.88 F1=0.82
- 1: AUC=0.86 F1=0.81
- 2: AUC=0.77 F1=0.7
- 3: AUC=0.65 F1=0.64
- 4: AUC=0.72 F1=0.64
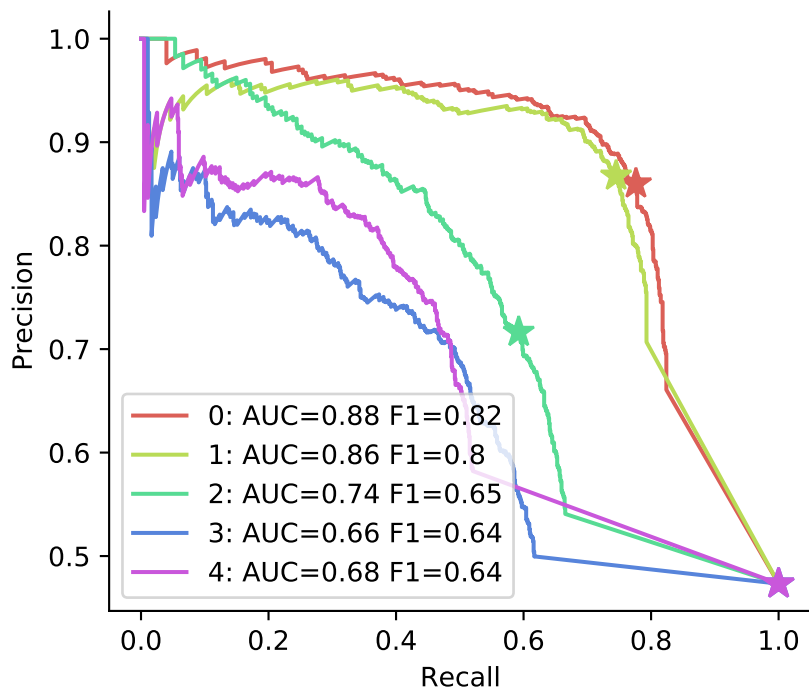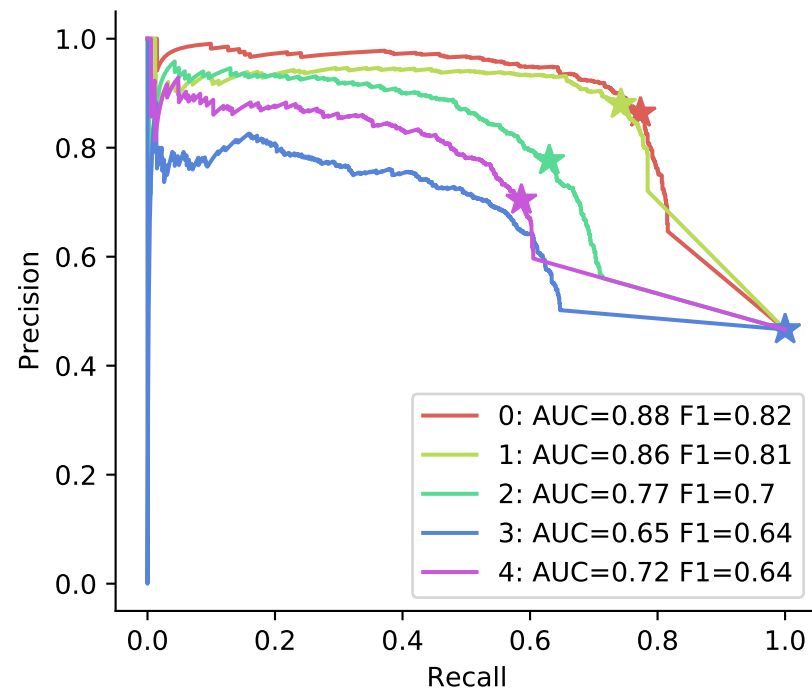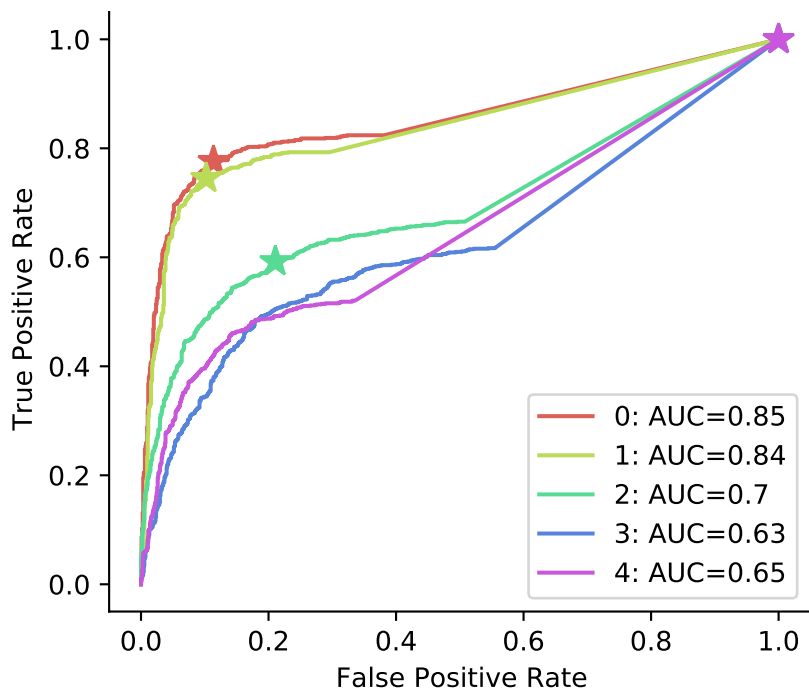
Model:final
construct_training_data.method
0: Substitute Term, Prepend Term
1: Prepend Term Only
2: Substitute Term Only
3: No Changes
4: Subsitute with Nonsense

valid (bottom left): True Positive Rate vs False Positive Rate
- 0: AUC=0.85
- 1: AUC=0.84
- 2: AUC=0.7
- 3: AUC=0.63
- 4: AUC=0.65

test (bottom right): True Positive Rate vs False Positive Rate
- 0: AUC=0.85
- 1: AUC=0.84
- 2: AUC=0.74
- 3: AUC=0.65
- 4: AUC=0.69