BiDAF:

EM: 60.108; F1: 63.626

ALBERT Fine-Tune:

1. albert-base-v2

{'exact': 77.44323790720632, 'f1': 79.6039861382432, 'total': 6078, 'HasAns\_exact': 69.65635738831615, 'HasAns\_f1': 74.16942534303864, 'HasAns\_total': 2910, 'NoAns\_exact': 84.5959595959596, 'NoAns\_f1': 84.5959595959596, 'NoAns\_total': 3168, 'best\_exact': 77.44323790720632, 'best\_exact\_thresh': 0.0, 'best\_f1': 79.60398613824324, 'best\_f1\_thresh': 0.0}

2. albert-large-v2

{'exact': 78.57847976307995, 'f1': 81.3588516631804, 'total': 6078, 'HasAns\_exact': 72.98969072164948, 'HasAns\_f1': 78.79694172124073, 'HasAns\_total': 2910, 'NoAns\_exact': 83.71212121212122, 'NoAns\_f1': 83.71212121212122, 'NoAns\_total': 3168, 'best\_exact': 78.57847976307995, 'best\_exact\_thresh': 0.0, 'best\_f1': 81.35885166318045, 'best\_f1\_thresh': 0.0}

ALBERT + BiDAF Ensemble:

Raw: EM=39.6, F1=43.0;

a. BiDAF Ensemble:

Classifier: Acc=60.9;

Ensemble: EM=53.5, F1=66.5;

b.

ALBERT Ensemble:

Use albert-base-v2

Classifier: F1=0.83, EM=0.82

Ensemble: F1=80.557, EM=78.496;

Use albert-large-v2

Classifier: F1=0.85, EM=0.85;

Ensemble: F1=81.959, EM=79.253;

Test Submission:

F1=82.533, EM=79.746