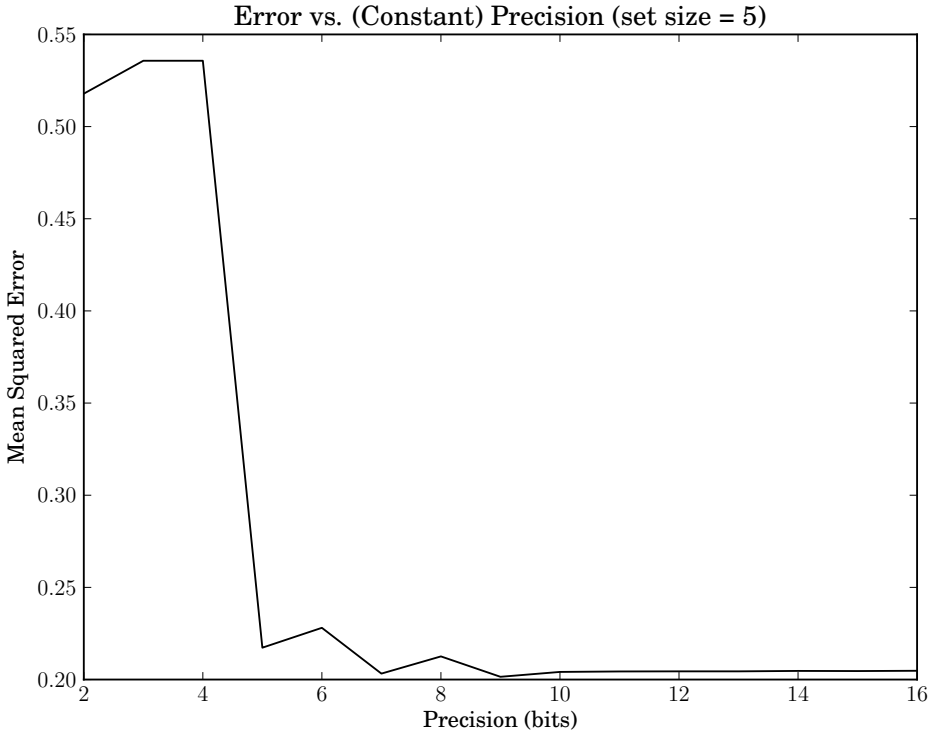
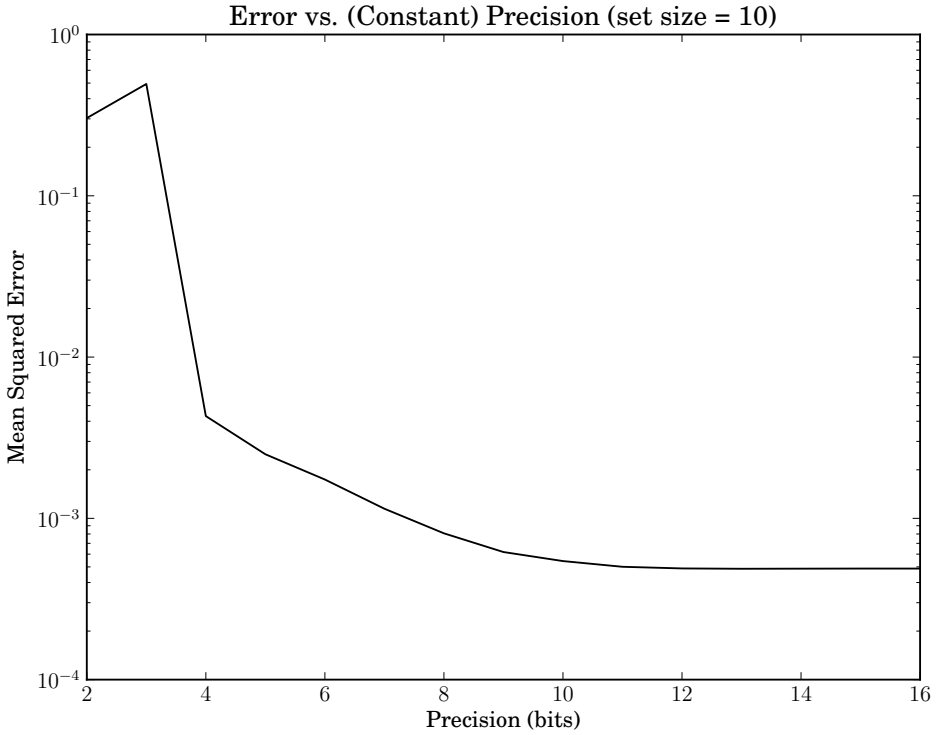
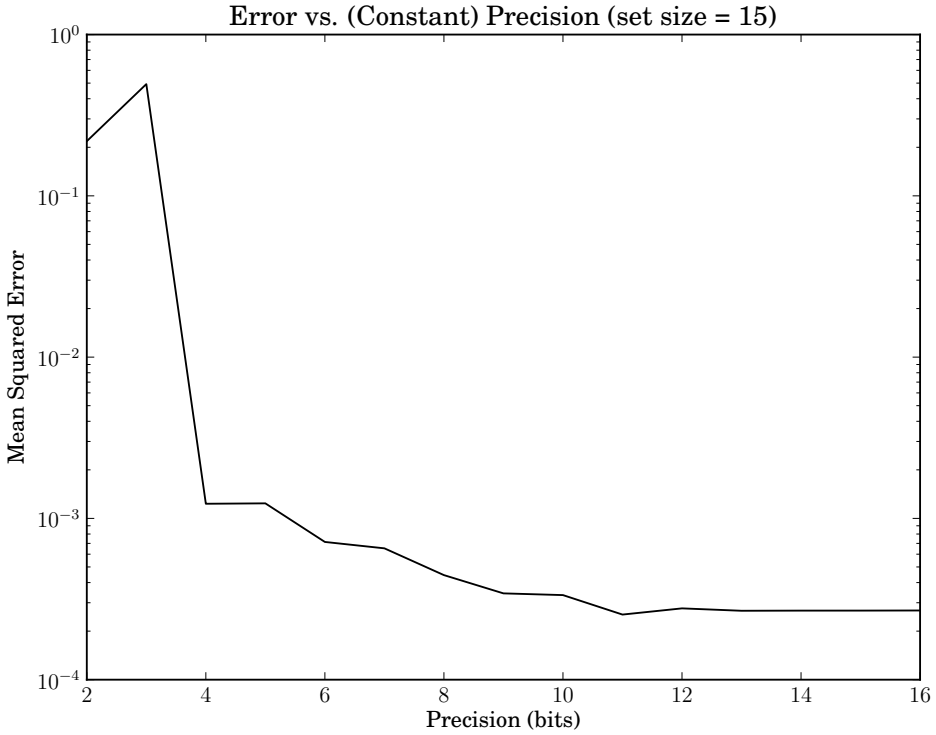
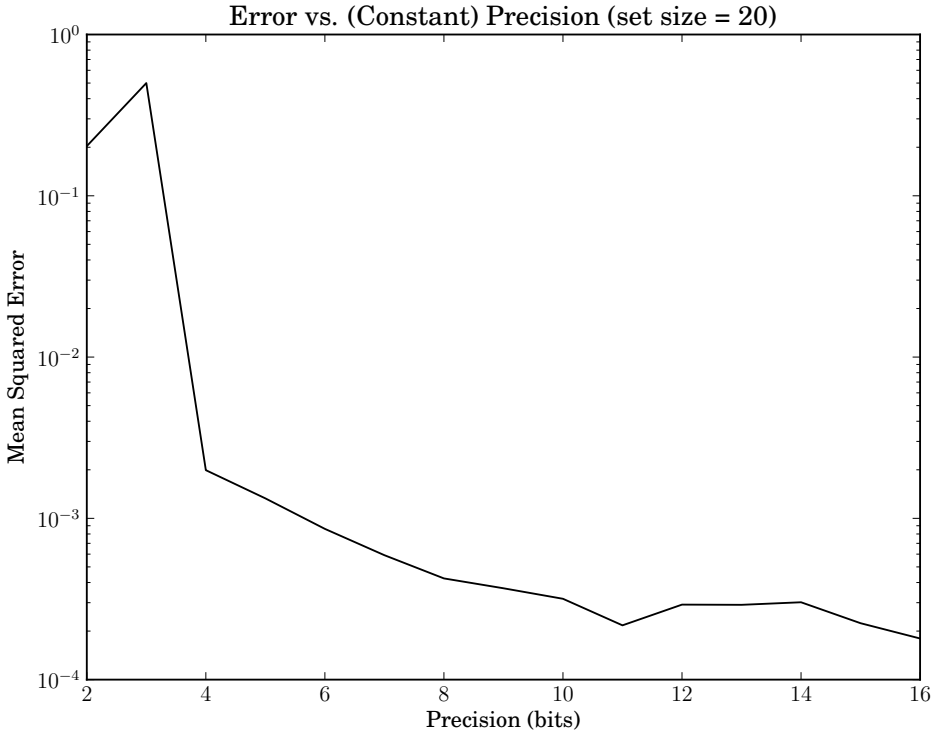
Mean squared error averaged over all testing data versus (constant) precision (in bits) of the network. Error is on a log scale. For a given set size s, network is trained on 2^{s-3} of the 2^s possible inputs (selected at random); tested on the remaining inputs. Learning rate $\eta = 4.0$, 400 iterations over training data. The first bit of each input represents the distance (bit in the string) that should be attended to (all 0 in these experiments). Hidden layer contains 20 neurons.

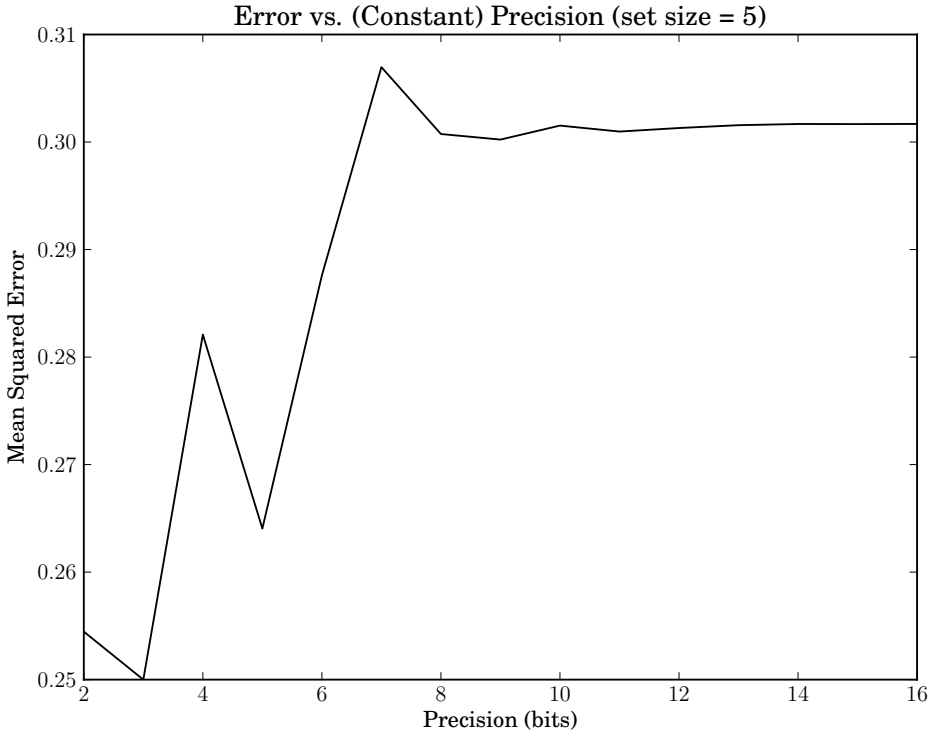


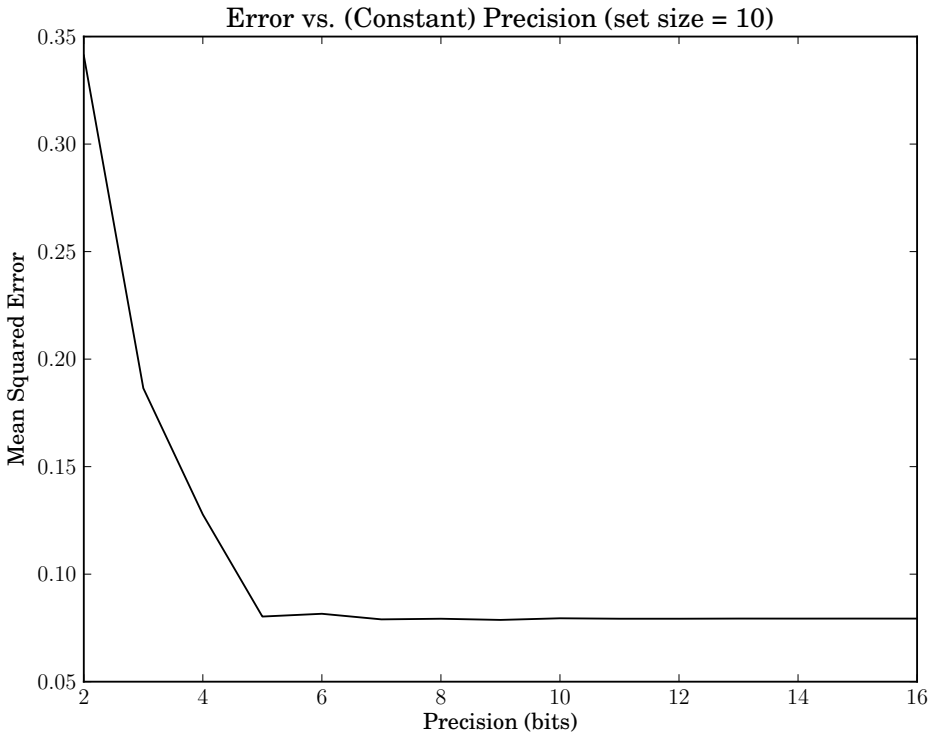

Error vs. (Constant) Precision (set size = 10)
Mean Squared Error
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
2
4
6
8
10
12
14
16
Precision (bits)

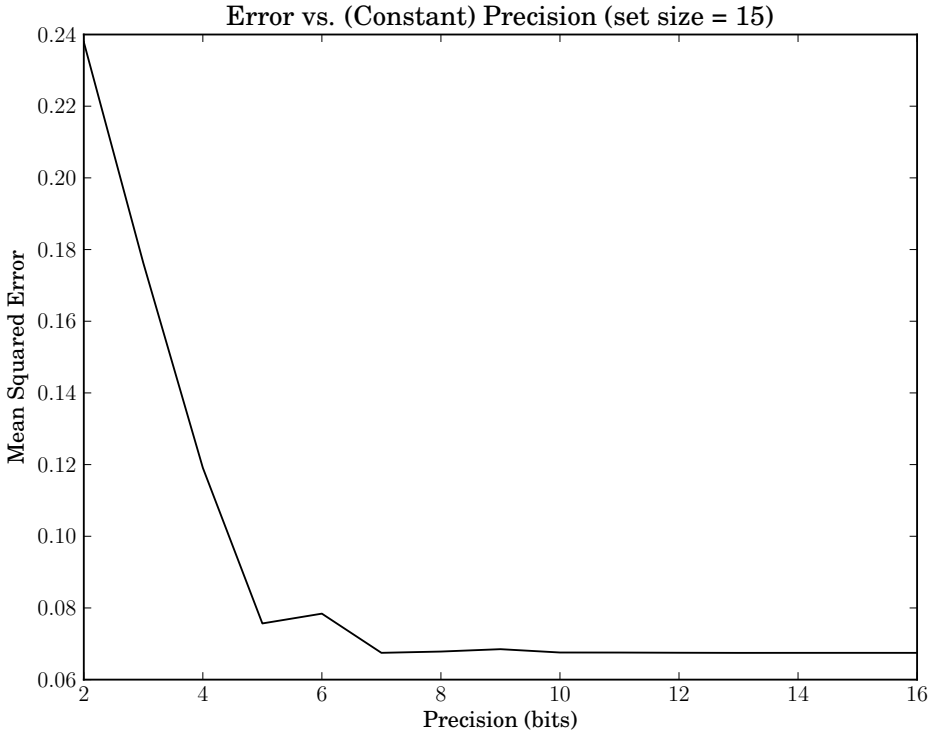

Error vs. (Constant) Precision (set size = 15)
Mean Squared Error
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
2
4
6
8
10
12
14
16
Precision (bits)



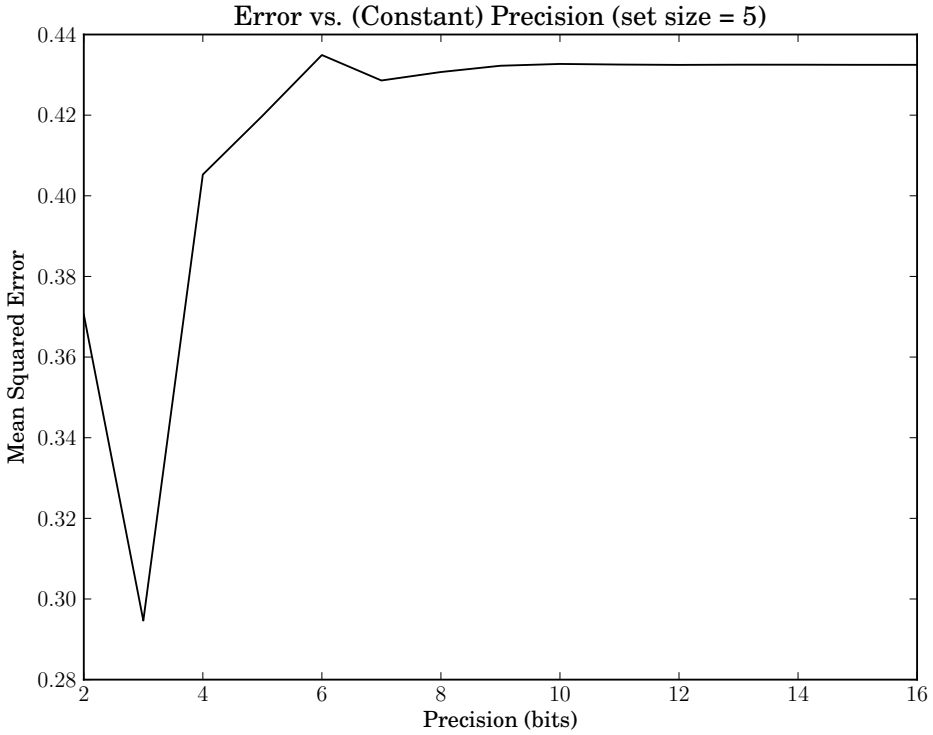
Same as above except: $\eta = 1.0$, 1 iteration over training data.

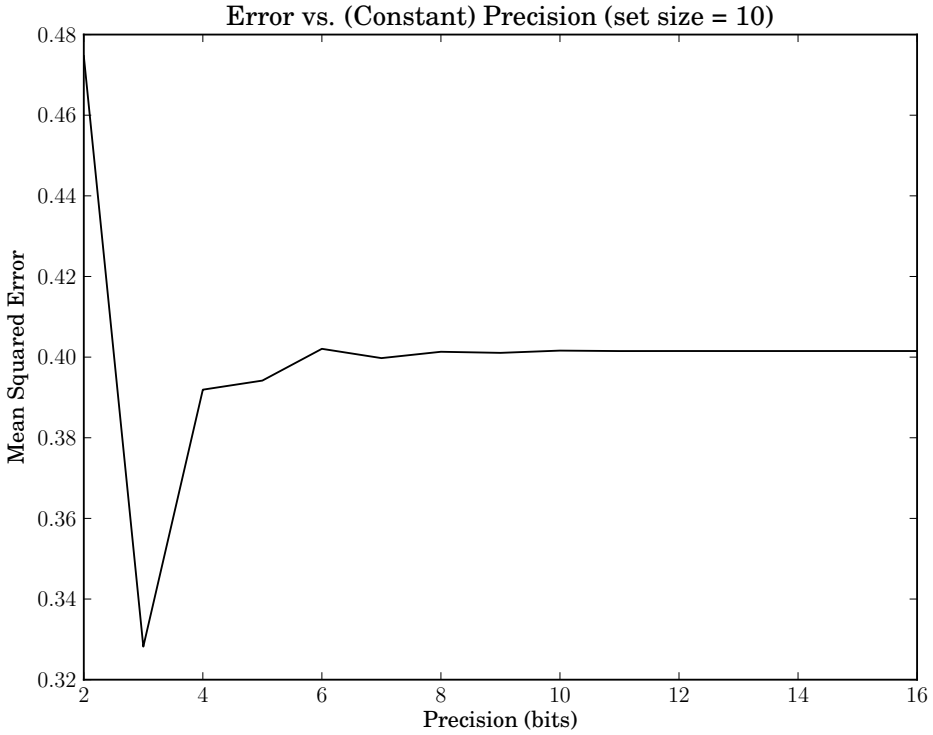

Error vs. (Constant) Precision (set size = 5)
Mean Squared Error
0.31
0.30
0.29
0.28
0.27
0.26
0.25
2
4
6
8
10
12
14
16
Precision (bits)

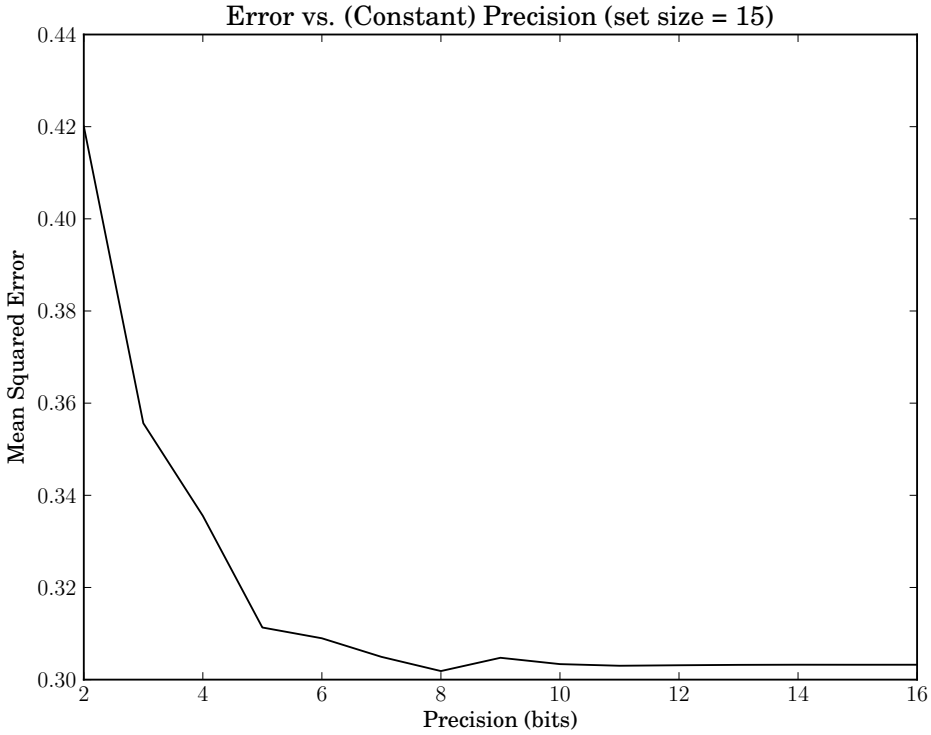

Error vs. (Constant) Precision (set size = 10)
Mean Squared Error
0.35
0.30
0.25
0.20
0.15
0.10
0.05
2
4
6
8
10
12
14
16
Precision (bits)



Same as above except: $\eta = 1.0$, 1 iteration over training data.


Error vs. (Constant) Precision (set size = 5)
Mean Squared Error
0.44
0.42
0.40
0.38
0.36
0.34
0.32
0.30
0.28
2
4
6
8
10
12
14
16
Precision (bits)


Error vs. (Constant) Precision (set size = 10)
Mean Squared Error
0.48
0.46
0.44
0.42
0.40
0.38
0.36
0.34
0.32
2
4
6
8
10
12
14
16
Precision (bits)


Error vs. (Constant) Precision (set size = 15)
Mean Squared Error
0.44
0.42
0.40
0.38
0.36
0.34
0.32
0.30
2
4
6
8
10
12
14
16
Precision (bits)