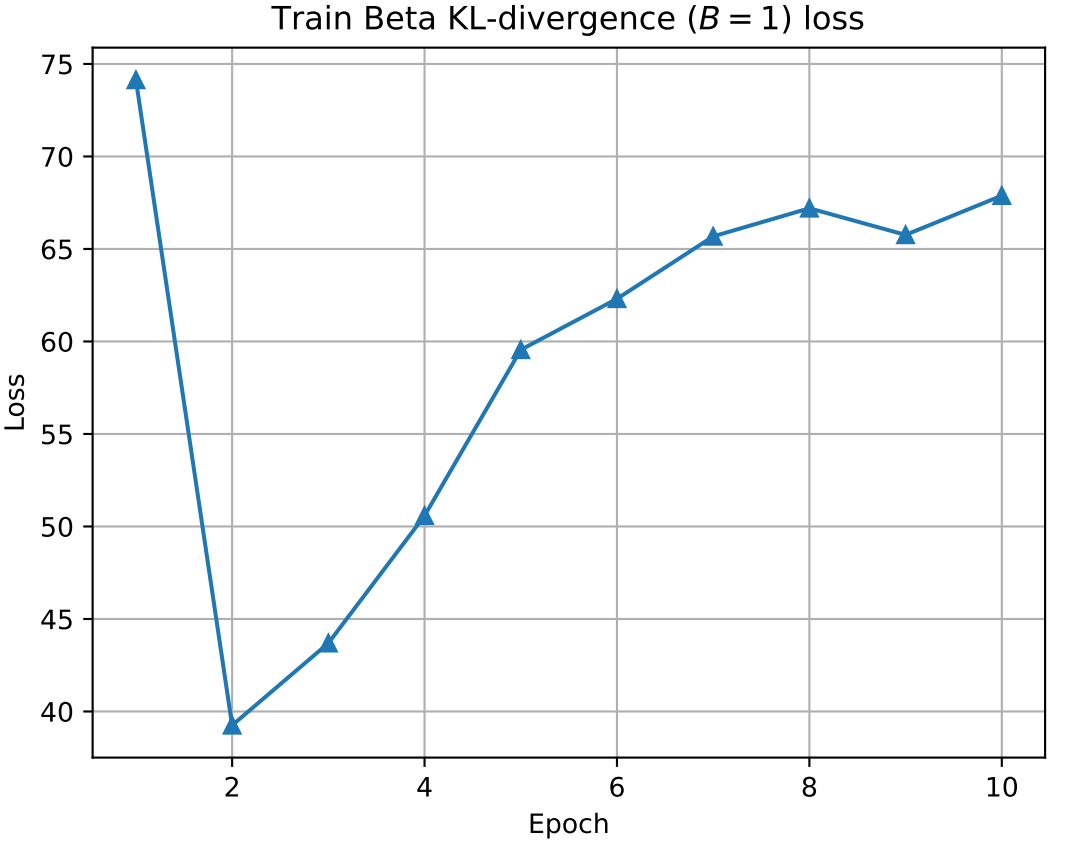
Train Beta KL-divergence ($B = 1$) loss
Loss
Epoch
75
70
65
60
55
50
45
40
2
4
6
8
10