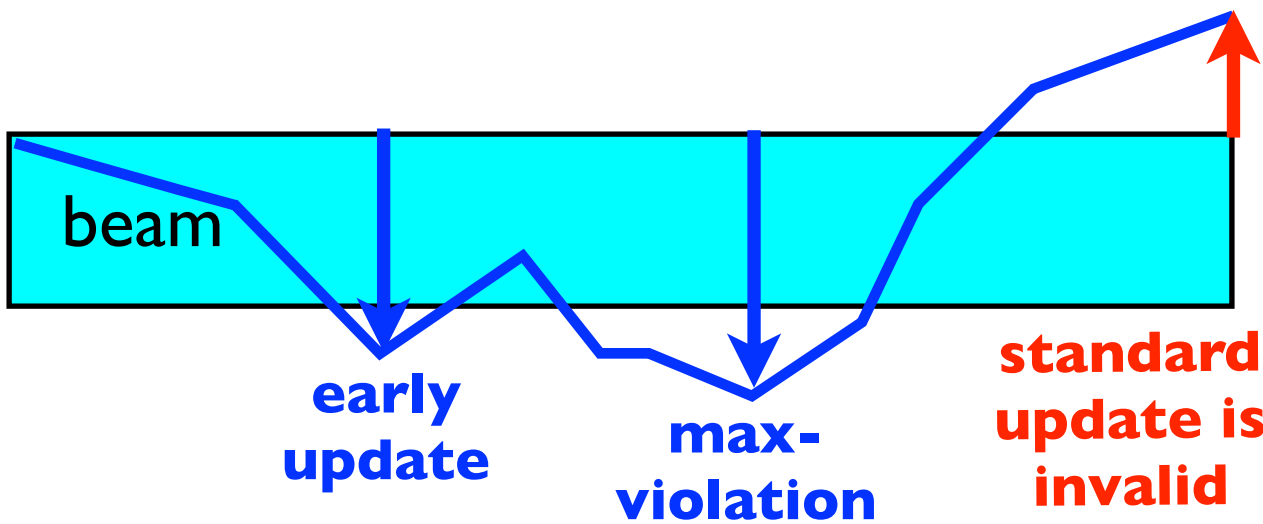


# Large-scale Discriminative Training

**Finally made successful for the first time**

Standard Perceptron (Liang et al 2006)

**doesn't work well due to search errors**



our work: violation-fixing perceptron  
with 20M+ features: +2 BLEU over MERT

training set (> 100k sentences)

MERT  
(Och '03)

~10 features

MIRA  
(Watanabe+ '07)  
(Chiang+ '08-'12)

~10k features

PRO  
(Hopkins+May '11)

HOLS  
(Flanigan+ '13)

sparse features  
as one feature

...

dev set  
(~1k sents)

test set  
(~1k sents)