



$$L \left( \begin{matrix} \hat{y}_{start}^{pred} & \hat{y}_{end}^{pred} & y_{start}^{true} & y_{end}^{true} \end{matrix} \right)$$

penalize by:  
distance  
of end < start

$$L(\text{discrete})$$

