



target value for
start position (t_s)

predicted value for
start position (p_s)

categorical
cross entropy
(l_{start})