

Decoder

$1 - p$

y_{j-1}^{oracle}

y_{j-1}^*

p

Logistic regression
classifier

...

$1 - p$

p

... s_{j-1} → GRU cell

...