

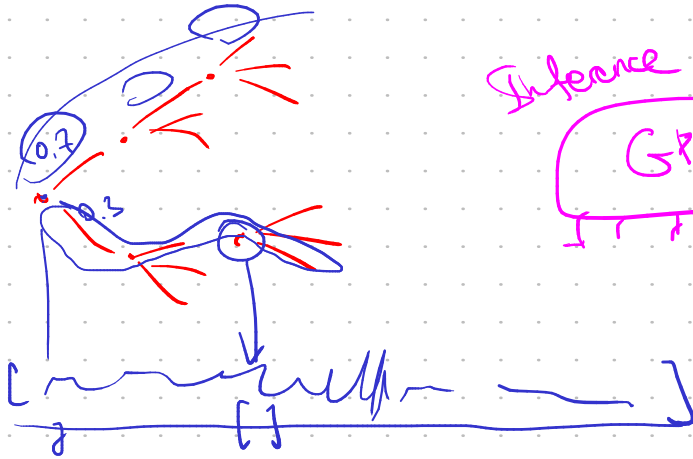
Perplexity

$x_1 - \dots - x_n$ - generated

$p(x_{t+1}) - \dots - p(x_n)$

Perplexity: $\left(\prod_{i=1}^n p(x_i) \right)^{-\frac{1}{N}}$ - reciprocal perplexity

$H(p)$ - bits per token
 perplex = $2^{H(p)}$



Training

