log_probs

[L, V] GeneralFunction log_softmax

logits

[L, V] Einsum ab,cb → ac

w.unembed

[V,d] Array

LayerNorm

$\frac{z-\bar{z}}{\sqrt{\frac{(z-\bar{z})^2}{d}+\epsilon}} * \gamma + \beta$

ln

[L, d] Add

ln.w.bias

[d] Array

β

$\frac{z-\bar{z}}{\sqrt{\frac{(z-\bar{z})^2}{d}+\epsilon}} * \gamma$

ln.y_scaled

[L, d] Einsum a,ab → ab

ln.w.scale

[d] Array

γ

$\frac{1}{\sqrt{\frac{(z-\bar{z})^2}{d}+\epsilon}}$

ln.rsqrt

[L,] GeneralFunction rsqrt

ln.y

[L, d] Einsum ab,b → ab

$\frac{z-\bar{z}}{\sqrt{\frac{(z-\bar{z})^2}{d}+\epsilon}}$

ϵ

ln.c.eps

[ ] Scalar 1e-5

ln.var_p_eps

[L,] Add

$\frac{(z-\bar{z})^2}{d}+\epsilon$

$\frac{1}{d}$

ln.c.recip_hidden_size

[L,] GeneralFunction reciprocal

(repeat)

ln.var

[L,] Einsum ,ab,ab → a

$\frac{(z-\bar{z})^2}{d}$

$z-\bar{z}$

ln.mean_subbed

[L, d] Einsum ab,bc → ac

ln.neg_mean

[L,1] Einsum Ah,z,A → Az

$-\bar{z}$

ln.neg

[1] Scalar -1

ln.c.recip_hidden_size

[L,] GeneralFunction reciprocal

$\frac{1}{d}$

z

ln.input

[L, d] Add

ln.c.hidden_size

[L,] GeneralFunction last_dim_size