Cross-entropy vs. negative-linear-output

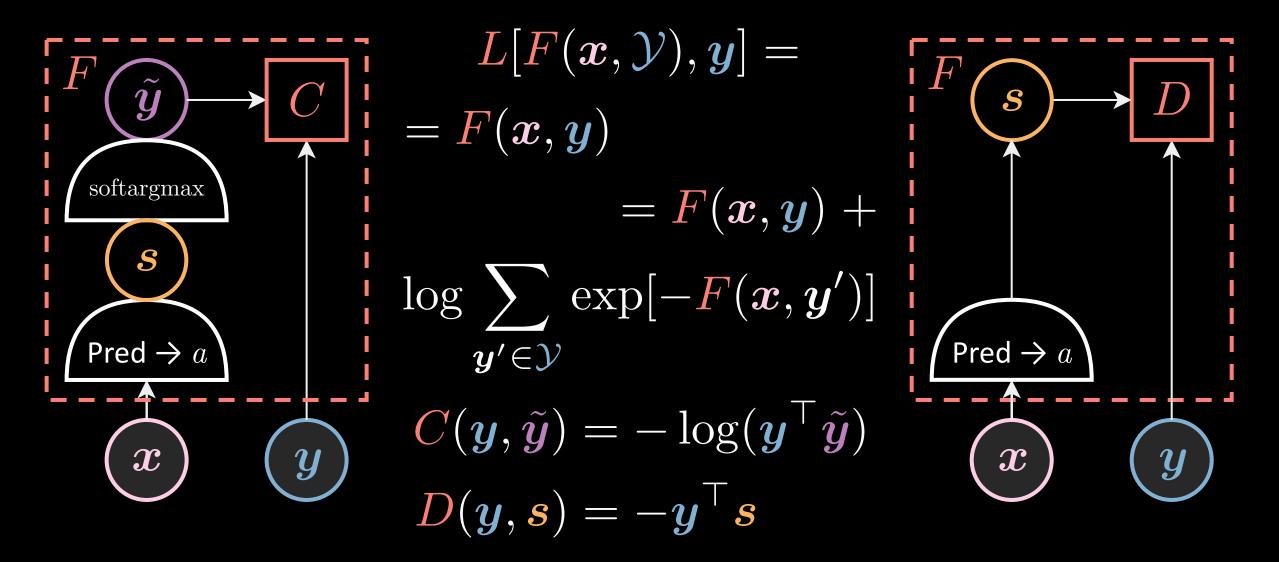
Two possible energy perspective for a classifier

$$F(oldsymbol{x},oldsymbol{y}) = C(ilde{oldsymbol{y}},oldsymbol{y})$$

 $F(oldsymbol{x},oldsymbol{y})=D(oldsymbol{y},oldsymbol{s})$

Cross-entropy

Negative linear output



Negative linear output energy $F(m{x},m{y}) = D(m{y},m{s})$

$$egin{aligned} D(oldsymbol{y}, oldsymbol{s}) &= -oldsymbol{y}^{ op} oldsymbol{s} \ D(oldsymbol{\mathcal{Y}}, oldsymbol{s}) &= -oldsymbol{s} & ext{negative} \ L[F(oldsymbol{x}, oldsymbol{\mathcal{Y}}), oldsymbol{y}] &= -rac{1}{eta} \log ig(oldsymbol{y}^{ op} \operatorname{softargmin}_{eta}[F(oldsymbol{x}, oldsymbol{\mathcal{Y}})] ig) & ext{loss functional} \ & ext{softargmin}_{eta}[F(oldsymbol{x}, oldsymbol{\mathcal{Y}})] &\doteq rac{\exp[-eta F(oldsymbol{x}, oldsymbol{\mathcal{Y}})]}{\sum_{oldsymbol{y}' \in \mathcal{Y}} \exp[-eta F(oldsymbol{x}, oldsymbol{y}')]} \end{aligned}$$

$$egin{aligned} oldsymbol{L}[oldsymbol{F}(oldsymbol{x}, oldsymbol{\mathcal{Y}}), oldsymbol{y}] &= oldsymbol{F}(oldsymbol{x}, oldsymbol{y}) + oldsymbol{egin{bmatrix} rac{1}{eta} \log \sum_{oldsymbol{y}' \in \mathcal{Y}} \exp[-eta oldsymbol{F}(oldsymbol{x}, oldsymbol{y}')] \ & -\operatorname{softmin}_{oldsymbol{y} \in \mathcal{Y}} [oldsymbol{F}(oldsymbol{x}, oldsymbol{y})]) \end{aligned} }$$

Gradients wrt the free energy $\hat{y} \mid x \in \mathcal{Y} \setminus y \mid x$

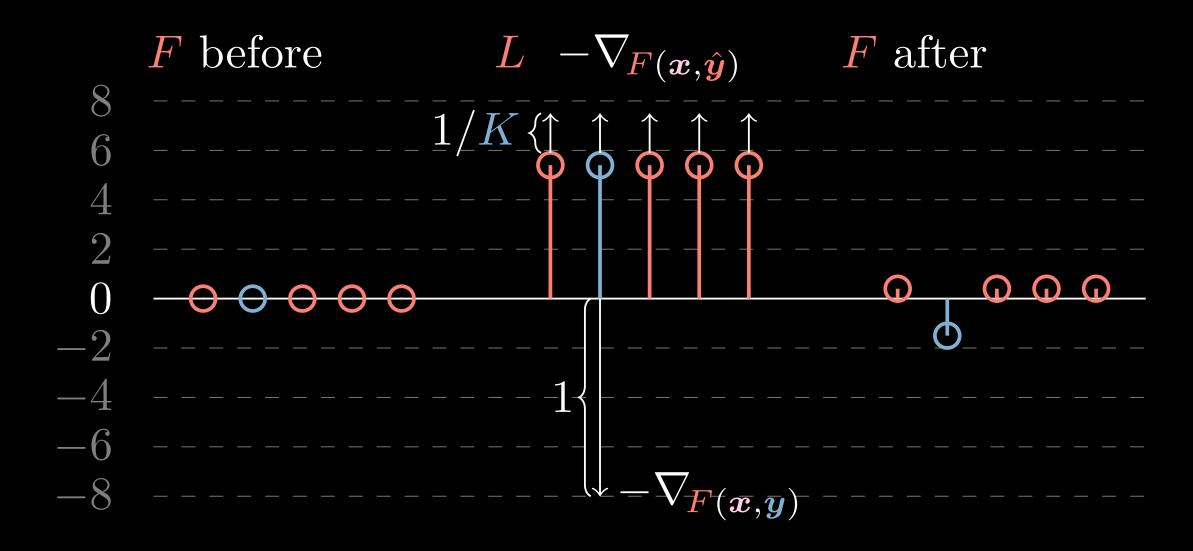
$$L[F(\boldsymbol{x}, \mathcal{Y}), \boldsymbol{y}] = F(\boldsymbol{x}, \boldsymbol{y}) - \operatorname{softmin}_{\beta}[F(\boldsymbol{x}, \mathcal{Y})]$$

$$egin{aligned}
abla_{m{F}(m{x},m{y})} m{L}[m{F}(m{x},m{\mathcal{Y}}),m{y}] &= 1 - rac{\exp[-eta m{F}(m{x},m{y})]}{\sum_{m{y}' \in m{\mathcal{Y}}} \exp[-eta m{F}(m{x},m{y}')]} = \ &= 1 - \mathbb{P}^{eta}_{m{w}}(m{y} \mid m{x}) = 1 - m{y}^{ op} m{ ilde{y}} \end{aligned}$$

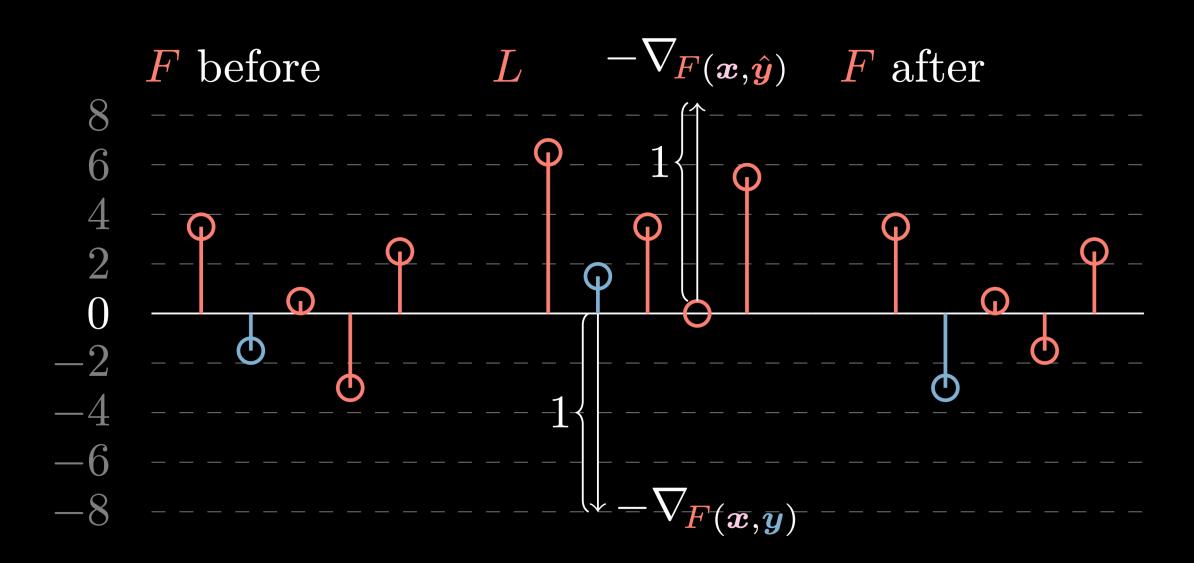
$$[
abla_{F(oldsymbol{x},\hat{oldsymbol{y}})}L[F(oldsymbol{x},\mathcal{Y}),oldsymbol{y}] = 0 - \mathbb{P}^{eta}_{oldsymbol{w}}(\hat{oldsymbol{y}} \mid oldsymbol{x}) = -\hat{oldsymbol{y}}^{ op} \hat{oldsymbol{y}}]$$

$$\mathbb{P}^{eta}_{m{w}}(m{y}'\midm{x}) \doteq \operatorname{softargmin}_{eta}[m{F}(m{x},m{\mathcal{Y}})]^{ op}m{y}' \qquad egin{array}{c}
abla_{m{F}(m{x},m{\mathcal{Y}})} L = \\
= m{y} - ilde{m{y}}
abla_{m{w}}(m{y}'\midm{x}) = \mathbf{y} - ilde{m{y}}
abla_{m{w}}(m{y}'\midm{y}) = \mathbf{y} - ilde{m{y}}
abla_{m{w}}(m{y}\midm{y}) = \mathbf{y} - \mathbf{y}$$

$$abla_{F(oldsymbol{x},\mathcal{Y})}L[F(oldsymbol{x},\mathcal{Y}),oldsymbol{y}] = oldsymbol{y} - ilde{oldsymbol{y}} \qquad F(oldsymbol{x},\mathcal{Y}) = oldsymbol{0}$$



$$abla_{F(oldsymbol{x},\mathcal{Y})}L[F(oldsymbol{x},\mathcal{Y}),oldsymbol{y}]=oldsymbol{y}- ilde{oldsymbol{y}} \qquad eta
ightarrow +\infty$$



$$abla_{F(oldsymbol{x},\mathcal{Y})}L[F(oldsymbol{x},\mathcal{Y}),oldsymbol{y}]=oldsymbol{y}- ilde{oldsymbol{y}}$$

