Cross-entropy *vs.* negative-linear-output

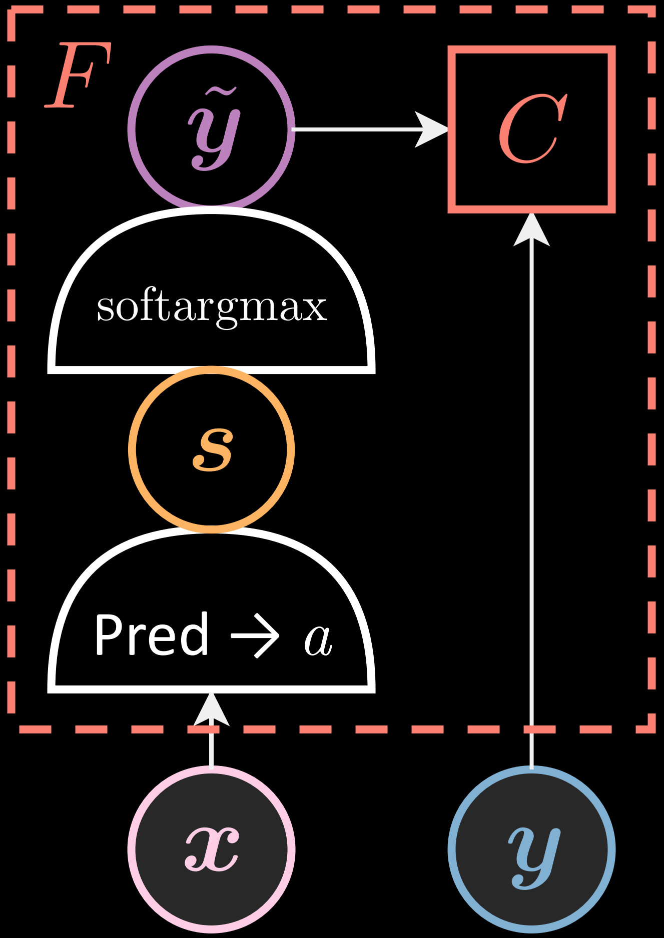
Two possible energy perspective for a classifier

$$F(\boldsymbol{x}, \boldsymbol{y}) = C(\tilde{\boldsymbol{y}}, \boldsymbol{y})$$

Cross-entropy

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

Negative linear output



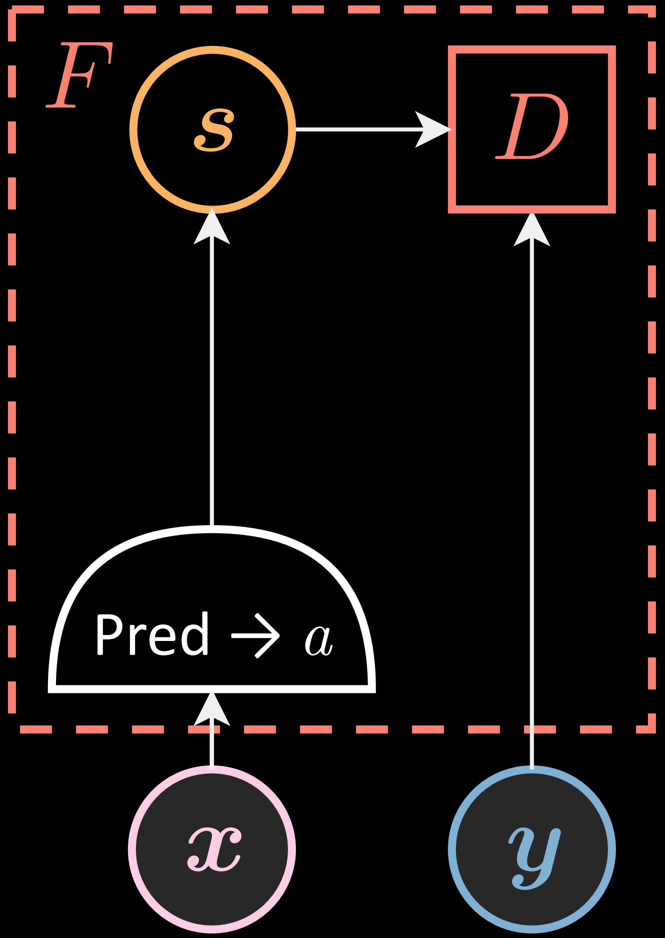
$$L[F(\boldsymbol{x}, \mathcal{Y}), \boldsymbol{y}] =$$

$$= F(\boldsymbol{x}, \boldsymbol{y})$$

$$= F(\boldsymbol{x}, \boldsymbol{y}) + \log \sum_{\boldsymbol{y}' \in \mathcal{Y}} \exp[-F(\boldsymbol{x}, \boldsymbol{y}')]$$

$$C(\boldsymbol{y}, \tilde{\boldsymbol{y}}) = -\log(\boldsymbol{y}^\top \tilde{\boldsymbol{y}})$$

$$D(\boldsymbol{y}, \boldsymbol{s}) = -\boldsymbol{y}^\top \boldsymbol{s}$$



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

$$D(\boldsymbol{y}, \boldsymbol{s}) = -\boldsymbol{y}^\top \boldsymbol{s} \qquad D(\mathcal{Y}, \boldsymbol{s}) = -\boldsymbol{s}$$

negative linear output

$$L[F(\boldsymbol{x}, \mathcal{Y}), \boldsymbol{y}] = -\frac{1}{\beta} \log\left(\boldsymbol{y}^\top \operatorname{softargmin}_\beta[F(\boldsymbol{x}, \mathcal{Y})]\right)$$

loss functional

$$\operatorname{softargmin}_\beta[F(\boldsymbol{x}, \mathcal{Y})] \doteq \frac{\exp[-\beta F(\boldsymbol{x}, \mathcal{Y})]}{\sum_{\boldsymbol{y}' \in \mathcal{Y}} \exp[-\beta F(\boldsymbol{x}, \boldsymbol{y}')]}$$

$$L[F(\boldsymbol{x}, \mathcal{Y}), \boldsymbol{y}] = F(\boldsymbol{x}, \boldsymbol{y}) + \boxed{\frac{1}{\beta} \log \sum_{\boldsymbol{y}' \in \mathcal{Y}} \exp[-\beta F(\boldsymbol{x}, \boldsymbol{y}')]}$$

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

Gradients wrt the free energy

$\hat{\boldsymbol{y}} \mid \boldsymbol{x} \in \mathcal{Y} \setminus \boldsymbol{y} \mid \boldsymbol{x}$

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

$$\nabla_{F(\boldsymbol{x},\boldsymbol{y})} L[F(\boldsymbol{x}, \mathcal{Y}), \boldsymbol{y}] = 1 - \frac{\exp[-\beta F(\boldsymbol{x}, \boldsymbol{y})]}{\sum_{\boldsymbol{y}' \in \mathcal{Y}} \exp[-\beta F(\boldsymbol{x}, \boldsymbol{y}')]} =$$

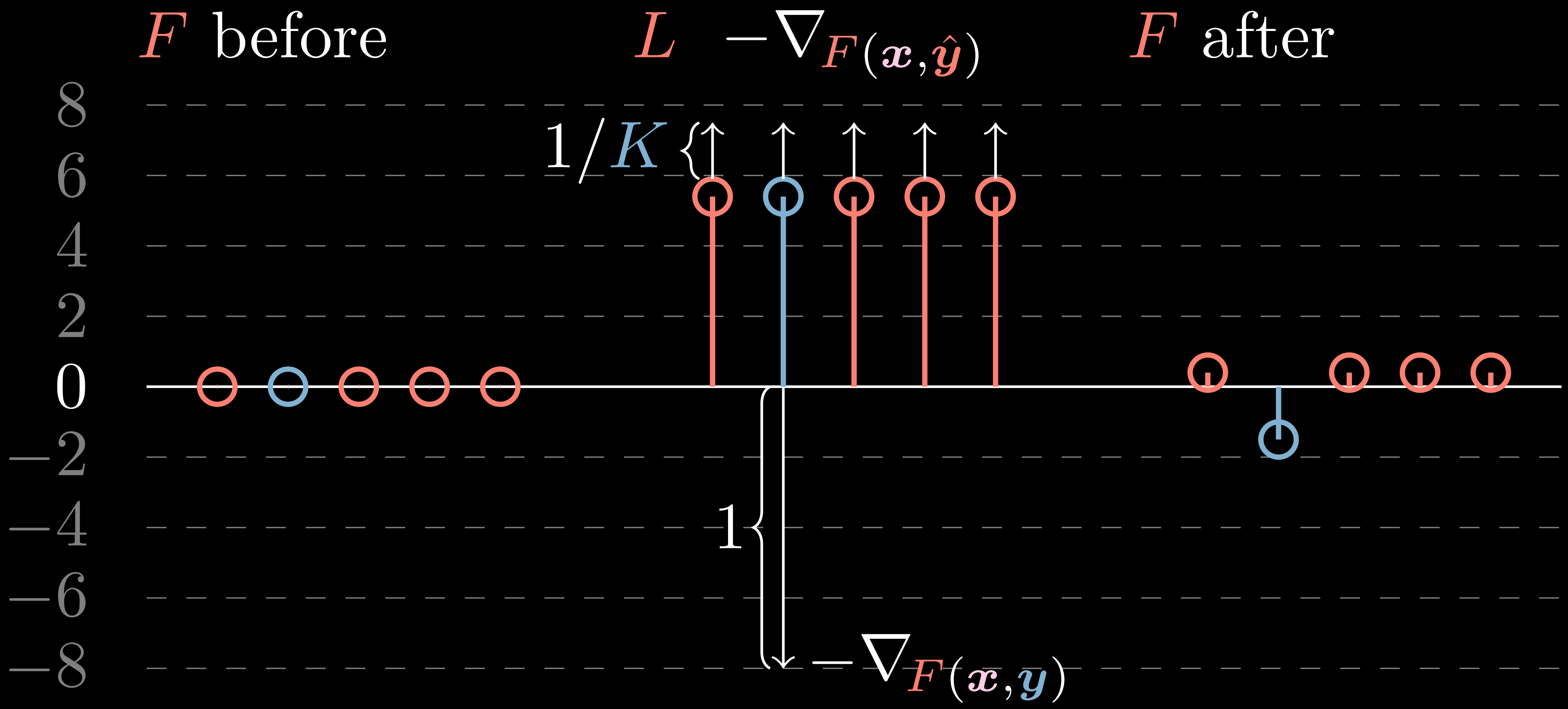
$$= 1 - \mathbb{P}_{\boldsymbol{w}}^{\beta}(\boldsymbol{y} \mid \boldsymbol{x}) = 1 - \boldsymbol{y}^\top \tilde{\boldsymbol{y}}$$

$$\nabla_{F(\boldsymbol{x},\hat{\boldsymbol{y}})} L[F(\boldsymbol{x}, \mathcal{Y}), \boldsymbol{y}] = 0 - \mathbb{P}_{\boldsymbol{w}}^{\beta}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}) = -\hat{\boldsymbol{y}}^\top \tilde{\boldsymbol{y}}$$

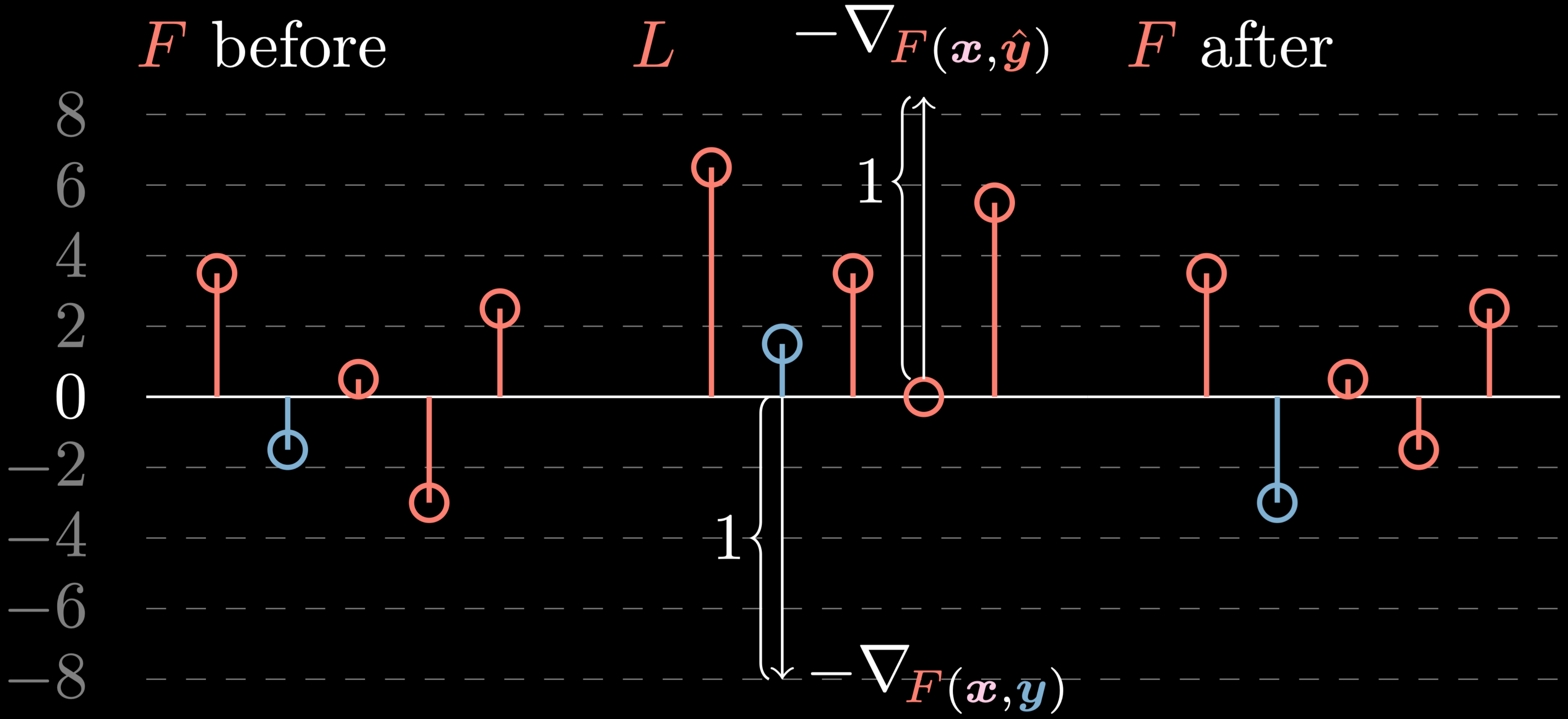
$$\mathbb{P}_{\boldsymbol{w}}^{\beta}(\boldsymbol{y}' \mid \boldsymbol{x}) \doteq \operatorname{softargmin}_\beta[F(\boldsymbol{x}, \mathcal{Y})]^\top \boldsymbol{y}'$$

$$\nabla_{F(\boldsymbol{x},\mathcal{Y})} L = \\ = \boldsymbol{y} - \tilde{\boldsymbol{y}}$$

$$\nabla_{F(\boldsymbol{x},\mathcal{Y})} L[F(\boldsymbol{x},\mathcal{Y}), \boldsymbol{y}] = \boldsymbol{y} - \tilde{\boldsymbol{y}} \qquad F(\boldsymbol{x},\mathcal{Y}) = \mathbf{0}$$



$$\nabla_{F(\boldsymbol{x},\mathcal{Y})} L[F(\boldsymbol{x},\mathcal{Y}),\boldsymbol{y}] = \boldsymbol{y} - \tilde{\boldsymbol{y}} \qquad \beta \to +\infty$$



$$\nabla_{F(\boldsymbol{x},\mathcal{Y})} L[F(\boldsymbol{x},\mathcal{Y}), \boldsymbol{y}] = \boldsymbol{y} - \tilde{\boldsymbol{y}}$$