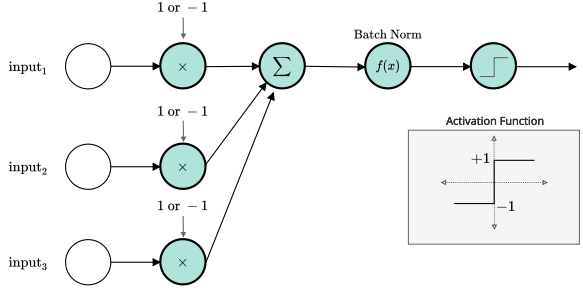
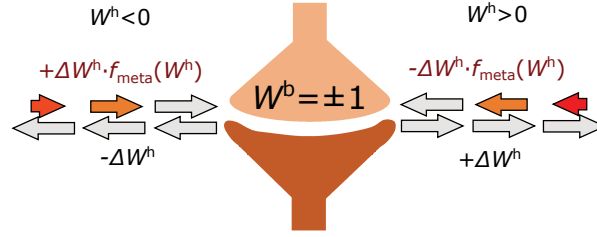
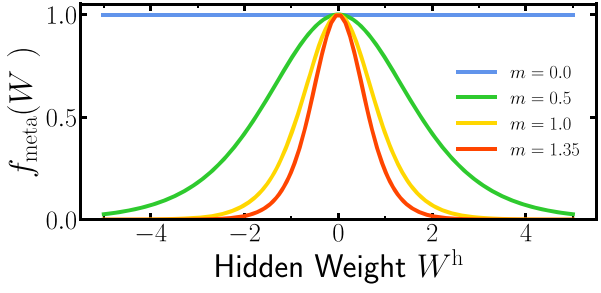


a



b



c

Input: $\mathbf{W^h}$, $\boldsymbol{\theta}^{\mathrm{BN}}$, $\mathbf{U_W}$, $\mathbf{U}_{\boldsymbol{\theta}}$, $(\mathbf{x}, \mathbf{y})$, m, η.
Output: $\mathbf{W^h}$, $\boldsymbol{\theta}^{\mathrm{BN}}$, $\mathbf{U_W}$, $\mathbf{U}_{\boldsymbol{\theta}}$.
$\mathbf{W^b} \leftarrow \mathrm{Sign}(\mathbf{W^h})$ ▷Computing binary weights
$\hat{\mathbf{y}}$, cache $\leftarrow$ Forward($\mathbf{x}, \mathbf{W^b}, \boldsymbol{\theta}^{\mathrm{BN}}$) ▷Perform inference
$C \leftarrow \mathrm{Cost}(\hat{\mathbf{y}}, \mathbf{y})$ ▷Compute mean loss over the batch
$(\partial_{\mathbf{W}} C, \partial_{\boldsymbol{\theta}} C) \leftarrow$ Backward($C, \hat{\mathbf{y}}, \mathbf{W^b}, \boldsymbol{\theta}^{\mathrm{BN}}$, cache)
▷Cost gradients
$(\mathbf{U_W}, \mathbf{U}_{\boldsymbol{\theta}}) \leftarrow \mathrm{Adam}(\partial_{\mathbf{W}} C, \partial_{\boldsymbol{\theta}} C, \mathbf{U_W}, \mathbf{U}_{\boldsymbol{\theta}})$
for W^{h} in $\mathbf{W^h}$ **do**
if $U_W \cdot W^{\mathrm{b}} > 0$ **then** ▷If U_W prescribes to decrease$|W^{\mathrm{b}}|$
$W^{\mathrm{h}} \leftarrow W^{\mathrm{h}} - \eta U_W \cdot f_{\mathrm{meta}}(m,\ W^{\mathrm{h}})$ ▷Metaplastic update
else
$W^{\mathrm{h}} \leftarrow W^{\mathrm{h}} - \eta U_W$
end if
end for
$\boldsymbol{\theta}^{\mathrm{BN}} \leftarrow \boldsymbol{\theta}^{\mathrm{BN}} - \eta \mathbf{U}_{\boldsymbol{\theta}}$
return $\mathbf{W^h}$, $\boldsymbol{\theta}^{\mathrm{BN}}$, $\mathbf{U_W}$, $\mathbf{U}_{\boldsymbol{\theta}}$

d