深度学习之 RNN

荣耀学院

2022 年 07 月

第一版序言

本书是《零基础自学神经网络 BP 算法》的续篇。熟悉后者可以根据本书快速掌握 RNN 神经网络的本质。考虑到熟悉后者，则具备一定的基础，因此本书只给出关键推导，次级细节不再详述。

本书最佳使用方式：拿出纸和笔，亲自推导所有公式。为了便于理解和学习，本书所有的推导给出了所有步骤，不做任何省略。

目 录

第一版序言 **iii**

第一章　最简 RNN **1**

第二章　最简 RNN 的 BP 算法 **3**

第三章　LSTM 的 BP 算法 **9**

第一章　最简 RNN

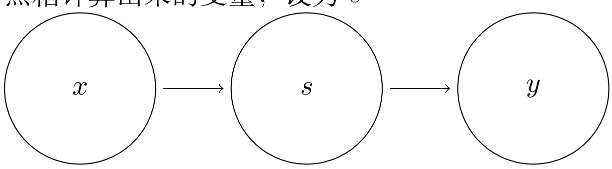
假如有一个分类器，这个分类器是黑箱：给黑箱输入 x，黑箱经过计算，输出 y。

这个黑箱是这样的：

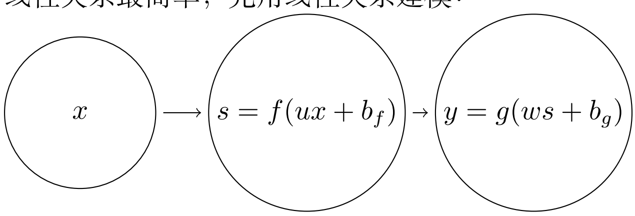
黑箱，不知道它是如何运行的，设计一个原理清晰的东西取代它。

直觉上，需要把黑箱拆成三层：输入层，处理层，输出层：

黑箱计算出来的变量，设为 s：



线性关系最简单，先用线性关系建模：



这里，$f(\cdot)$ 和 $g(\cdot)$ 是激活函数，比如 sigmod 函数，其中,u 和 w 是常数，由黑箱决定的常数，不同的黑箱 u 和 w 是不同的，但对确定的黑箱来说，u 和 w 是固定的。

看起来好象差不多了。

经过研究发现，有一个巨大的变动，这个黑箱计算输出值 s 的时候，会使用上一个时刻的 s，也就是说，如果 t 是时间点，计算 s_{t+1} 的时候，需要使用 s_t 的值，但不清楚具体系数。

那模型就要改动了，加上时间：

$$x_{t+1} \longrightarrow s = f(ux_{t+1} + vs_t + b_f) \rightarrow y = g(ws_{t+1} + b_g)$$

其中，v 也是黑箱决定的常数，跟 u 和 w 的性质是一样的。

这就是最简单的 RNN。

如果把这个 RNN 扩展成深度 RNN，那就是处理层有多个头尾巴串联的神经元，且数量足够多。

第二章　最简 RNN 的 BP 算法

最简 RNN 有五个参数：u，v，w，b_f，b_g。

用 BP 算法求解。

假设有一对 $t+1$ 时刻的训练样本，$[x_{t+1}, y_{t+1}]$，x_{t+1} 是输入，y_{t+1} 是目标值，t 时刻的 s 值 s_t 是已知的，RNN 的输出是 o_{t+1}。训练的目标，是让 o_{t+1} 尽可能接近 y_{t+1}，其中 $t = 0, ..., T-1$。

定义目标函数：

$$Err = \sum_{t=0}^{T-1} \frac{1}{2}(o_{t+1} - y_{t+1})^2$$

根据微分求极值，迭代计算 u，v，w，使得 Err 最小，迭代公式是：

$$u = u - \eta \frac{\partial Err}{\partial u}$$

$$v = v - \eta \frac{\partial Err}{\partial v}$$

$$w = w - \eta \frac{\partial Err}{\partial w}$$

其中，$\eta \in (0,1)$，是学习速率。

推导 Err 对 u，v，w，b_f 和 b_g 的一阶偏导：

$$
\begin{aligned}
\frac{\partial Err}{\partial w} &= \sum_{t=0}^{T-1} \frac{\partial(\frac{1}{2}(o_{t+1}-y_{t+1})^2)}{\partial w} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) \frac{\partial o_{t+1}}{\partial w} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) \frac{\partial g(ws_{t+1}+b_g)}{\partial w} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) g^{'}(ws_{t+1}+b_g) \frac{\partial (ws_{t+1}+b_g)}{\partial w} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) g^{'}(ws_{t+1}+b_g) s_{t+1}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Err}{\partial b_g} &= \sum_{t=0}^{T-1} \frac{\partial(\frac{1}{2}(o_{t+1}-y_{t+1})^2)}{\partial b_g} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) \frac{\partial o_{t+1}}{\partial b_g} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) \frac{\partial g(ws_{t+1}+b_g)}{\partial b_g} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) g^{'}(ws_{t+1}+b_g) \frac{\partial (ws_{t+1}+b_g)}{\partial b_g} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) g^{'}(ws_{t+1}+b_g)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Err}{\partial v} &= \sum_{t=0}^{T-1} \frac{\partial(\frac{1}{2}(o_{t+1}-y_{t+1})^2)}{\partial v} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial o_{t+1}}{\partial v} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial g(ws_{t+1}+b_g)}{\partial v} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)\frac{\partial(ws_{t+1}+b_g)}{\partial v} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)w\frac{\partial s_{t+1}}{\partial v} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)w\frac{\partial f(ux_{t+1}+vs_t+b_f)}{\partial v} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)wf^{'}(ux_{t+1}+vs_t+b_f)\frac{\partial(ux_{t+1}+vs_t+b_f)}{\partial v} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)wf^{'}(ux_{t+1}+vs_t+b_f)s_t
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Err}{\partial u} &= \sum_{t=0}^{T-1} \frac{\partial(\frac{1}{2}(o_{t+1}-y_{t+1})^2)}{\partial u} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial o_{t+1}}{\partial u} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial g(ws_{t+1}+b_g)}{\partial u} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)\frac{\partial(ws_{t+1}+b_g)}{\partial u} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)w\frac{\partial s_{t+1}}{\partial u} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)w\frac{\partial f(ux_{t+1}+vs_t+b_f)}{\partial u} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)wf^{'}(ux_{t+1}+vs_t+b_f)\frac{\partial(ux_{t+1}+vs_t+b_f)}{\partial u} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)wf^{'}(ux_{t+1}+vs_t+b_f)x_{t+1}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Err}{\partial b_f} &= \sum_{t=0}^{T-1}\frac{\partial(\frac{1}{2}(o_{t+1}-y_{t+1})^2)}{\partial b_f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial o_{t+1}}{\partial b_f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial g(ws_{t+1}+b_g)}{\partial b_f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)\frac{\partial(ws_{t+1}+b_g)}{\partial b_f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)w\frac{\partial s_{t+1}}{\partial b_f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)w\frac{\partial f(ux_{t+1}+vs_t+b_f)}{\partial b_f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)wf^{'}(ux_{t+1}+vs_t+b_f)\frac{\partial(ux_{t+1}+vs_t+b_f)}{\partial b_f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})g^{'}(ws_{t+1}+b_g)wf^{'}(ux_{t+1}+vs_t+b_f)
\end{aligned}
$$

RNN 的核心部分就这些了，剩下的就是编程实现问题，可以参考《零基础自学神经网络 BP 算法》，不复杂。

第三章　LSTM 的 BP 算法

LSTM: long short term memory。

x_t 是 t 时刻的输入，x_{t+1} 是 $t+1$ 时刻的输入，o_{t+1} 是 $t+1$ 时刻的输出。y_{t+1} 是 $t+1$ 时刻的目标值。

$\sigma(\cdot)$ 是 sigma 激活函数，$tanh(\cdot)$ 是 tanh 函数。

假设 t 时刻的所有量都计算完成了，所有量都是已知的。

LSTM 的输出，即按照如下次序计算 o_{t+1}：

遗忘门 (forget gate) 的输出：

$$f_{t+1} = \sigma(u^f o_t + v^f x_{t+1} + b^f)$$

输入门 (input gate) 的输出：

$$i_{t+1} = \sigma(u^i o_t + v^i x_{t+1} + b^i)$$

$$\widetilde{C}_{t+1} = tanh(u^C o_t + v^C x_{t+1} + b^C)$$

$$C_{t+1} = f_{t+1} C_t + i_{t+1} \widetilde{C}_{t+1}$$

输出门 (output gate) 的输出

$$h_{t+1} = \sigma(u^o o_t + v^o x_{t+1} + b^o)$$

$$o_{t+1} = tanh(C_{t+1}) h_{t+1}$$

其中，u^f，v^f，b^f，u^i，v^i，b^i，u^C，v^C，b^C，u^o，v^o，b^o 是 LSTM 的参数。f_{t+1} 是 $t+1$ 时刻遗忘门的输出。i_{t+1} 是 $t+1$ 时刻输入门的输出。o_{t+1} 是 $t+1$ 时刻输出门的输出，同时也是 LSTM 的输出。

尽管公式很多，实际上不复杂，推导只是个体力活。

目标函数：

$$Err = \sum_{t=0}^{T-1} \frac{1}{2}(o_{t+1} - y_{t+1})^2$$

推导 Err 对 u^f，v^f，b^f，u^i，v^i，b^i，u^C，v^C，b^C，u^o，v^o，b^o 的一阶偏导。

$$\begin{aligned}
\frac{\partial Err}{\partial u^o} &= \sum_{t=0}^{T-1} \frac{\partial \frac{1}{2}(o_{t+1} - y_{t+1})^2}{\partial u^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial (o_{t+1} - y_{t+1})}{\partial u^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial o_{t+1}}{\partial u^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial (tanh(C_{t+1}) h_{t+1})}{\partial u^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) tanh(C_{t+1}) \frac{\partial h_{t+1}}{\partial u^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) tanh(C_{t+1}) \sigma^{'}(u^o o_t + v^o x_{t+1} + b^o) o_t
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial Err}{\partial v^o} &= \sum_{t=0}^{T-1} \frac{\partial \frac{1}{2}(o_{t+1} - y_{t+1})^2}{\partial v^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial (o_{t+1} - y_{t+1})}{\partial v^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial o_{t+1}}{\partial v^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial (tanh(C_{t+1}) h_{t+1})}{\partial v^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) tanh(C_{t+1}) \frac{\partial h_{t+1}}{\partial v^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) tanh(C_{t+1}) \sigma^{'} (u^o o_t + v^o x_{t+1} + b^o) x_{t+1}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Err}{\partial b^o} &= \sum_{t=0}^{T-1} \frac{\partial \frac{1}{2}(o_{t+1} - y_{t+1})^2}{\partial b^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial (o_{t+1} - y_{t+1})}{\partial b^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial o_{t+1}}{\partial b^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial (tanh(C_{t+1}) h_{t+1})}{\partial b^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) tanh(C_{t+1}) \frac{\partial h_{t+1}}{\partial b^o} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) tanh(C_{t+1}) \sigma^{'} (u^o o_t + v^o x_{t+1} + b^o)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Err}{\partial u^C} &= \sum_{t=0}^{T-1} \frac{\partial \frac{1}{2}(o_{t+1}-y_{t+1})^2}{\partial u^C} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) \frac{\partial (o_{t+1}-y_{t+1})}{\partial u^C} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) \frac{\partial o_{t+1}}{\partial u^C} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) \frac{\partial (tanh(C_{t+1})h_{t+1})}{\partial u^C} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) h_{t+1} \frac{\partial tanh(C_{t+1})}{\partial u^C} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \frac{\partial C_{t+1}}{\partial u^C} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \frac{\partial (f_{t+1}C_t + i_{t+1}\widetilde{C}_{t+1})}{\partial u^C} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) i_{t+1} \frac{\partial \widetilde{C}_{t+1}}{\partial u^C} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) i_{t+1} \frac{\partial tanh(u^C o_t + v^C x_{t+1} + b^C)}{\partial u^C} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) i_{t+1} tanh^{'}(u^C o_t + v^C x_{t+1} + b^C) \cdot \\
&\quad \frac{\partial (u^C o_t + v^C x_{t+1} + b^C)}{\partial u^C} \\
&= \sum_{t=0}^{T-1} (o_{t+1}-y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) i_{t+1} tanh^{'}(u^C o_t + v^C x_{t+1} + b^C) o_t
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Err}{\partial v^C} &= \frac{\partial \frac{1}{2}(o_{t+1} - y_{t+1})^2}{\partial v^C} \\
&= (o_{t+1} - y_{t+1})\frac{\partial (o_{t+1} - y_{t+1})}{\partial v^C} \\
&= (o_{t+1} - y_{t+1})\frac{\partial o_{t+1}}{\partial v^C} \\
&= (o_{t+1} - y_{t+1})\frac{\partial (tanh(C_{t+1})h_{t+1})}{\partial v^C} \\
&= (o_{t+1} - y_{t+1})h_{t+1}\frac{\partial tanh(C_{t+1})}{\partial v^C} \\
&= (o_{t+1} - y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\frac{\partial C_{t+1}}{\partial v^C} \\
&= (o_{t+1} - y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\frac{\partial (f_{t+1}C_t + i_{t+1}\widetilde{C}_{t+1})}{\partial v^C} \\
&= (o_{t+1} - y_{t+1})h_{t+1}tanh^{'}(C_{t+1})i_{t+1}\frac{\partial \widetilde{C}_{t+1}}{\partial v^C} \\
&= (o_{t+1} - y_{t+1})h_{t+1}tanh^{'}(C_{t+1})i_{t+1}\frac{\partial tanh(u^C o_t + v^C x_{t+1} + b^C)}{\partial v^C} \\
&= (o_{t+1} - y_{t+1})h_{t+1}tanh^{'}(C_{t+1})i_{t+1}tanh^{'}(u^C o_t + v^C x_{t+1} + b^C) \cdot \\
&\quad \frac{\partial (u^C o_t + v^C x_{t+1} + b^C)}{\partial v^C} \\
&= (o_{t+1} - y_{t+1})h_{t+1}tanh^{'}(C_{t+1})i_{t+1}tanh^{'}(u^C o_t + v^C x_{t+1} + b^C)x_{t+1}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Err}{\partial b^C} &= \sum_{t=0}^{T-1}\frac{\partial\frac{1}{2}(o_{t+1}-y_{t+1})^2}{\partial b^C}\\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial(o_{t+1}-y_{t+1})}{\partial b^C}\\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial o_{t+1}}{\partial b^C}\\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial(tanh(C_{t+1})h_{t+1})}{\partial b^C}\\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}\frac{\partial tanh(C_{t+1})}{\partial b^C}\\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\frac{\partial C_{t+1}}{\partial b^C}\\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\frac{\partial(f_{t+1}C_t+i_{t+1}\widetilde{C}_{t+1})}{\partial b^C}\\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})i_{t+1}\frac{\partial\widetilde{C}_{t+1}}{\partial b^C}\\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})i_{t+1}\frac{\partial tanh(u^Co_t+v^Cx_{t+1}+b^C)}{\partial b^C}\\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})i_{t+1}tanh^{'}(u^Co_t+v^Cx_{t+1}+b^C)\cdot\\
&\quad \frac{\partial(u^Co_t+v^Cx_{t+1}+b^C)}{\partial b^C}\\
&= (o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})i_{t+1}tanh^{'}(u^Co_t+v^Cx_{t+1}+b^C)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Err}{\partial u^i} &= \sum_{t=0}^{T-1} \frac{\partial \frac{1}{2}(o_{t+1} - y_{t+1})^2}{\partial u^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial (o_{t+1} - y_{t+1})}{\partial u^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial o_{t+1}}{\partial u^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial (tanh(C_{t+1}) h_{t+1})}{\partial u^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} \frac{\partial tanh(C_{t+1})}{\partial u^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \frac{\partial C_{t+1}}{\partial u^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \frac{\partial (f_{t+1} C_t + i_{t+1} \widetilde{C}_{t+1})}{\partial u^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \widetilde{C}_{t+1} \frac{\partial i_{t+1}}{\partial u^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \widetilde{C}_{t+1} \frac{\partial \sigma(u^i o_t + v^i x_{t+1} + b^i)}{\partial u^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \widetilde{C}_{t+1} \sigma^{'}(u^i o_t + v^i x_{t+1} + b^i) \frac{\partial (u^i o_t + v^i x_{t+1} + b^i)}{\partial u^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \widetilde{C}_{t+1} \sigma^{'}(u^i o_t + v^i x_{t+1} + b^i) o_t
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Err}{\partial v^i} &= \sum_{t=0}^{T-1} \frac{\partial \frac{1}{2}(o_{t+1} - y_{t+1})^2}{\partial v^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial (o_{t+1} - y_{t+1})}{\partial v^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial o_{t+1}}{\partial v^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial (tanh(C_{t+1}) h_{t+1})}{\partial v^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} \frac{\partial tanh(C_{t+1})}{\partial v^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \frac{\partial C_{t+1}}{\partial v^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \frac{\partial (f_{t+1} C_t + i_{t+1} \widetilde{C}_{t+1})}{\partial v^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \widetilde{C}_{t+1} \frac{\partial i_{t+1}}{\partial v^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \widetilde{C}_{t+1} \frac{\partial \sigma(u^i o_t + v^i x_{t+1} + b^i)}{\partial v^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \widetilde{C}_{t+1} \sigma^{'}(u^i o_t + v^i x_{t+1} + b^i) \frac{\partial (u^i o_t + v^i x_{t+1} + b^i)}{\partial v^i} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \widetilde{C}_{t+1} \sigma^{'}(u^i o_t + v^i x_{t+1} + b^i) x_{t+1}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Err}{\partial b^i} &= \sum_{t=0}^{T-1} \frac{\partial \frac{1}{2}(o_{t+1}-y_{t+1})^2}{\partial b^i} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial (o_{t+1}-y_{t+1})}{\partial b^i} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial o_{t+1}}{\partial b^i} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial (tanh(C_{t+1})h_{t+1})}{\partial b^i} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}\frac{\partial tanh(C_{t+1})}{\partial b^i} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\frac{\partial C_{t+1}}{\partial b^i} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\frac{\partial (f_{t+1}C_t+i_{t+1}\widetilde{C}_{t+1})}{\partial b^i} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\widetilde{C}_{t+1}\frac{\partial i_{t+1}}{\partial b^i} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\widetilde{C}_{t+1}\frac{\partial \sigma(u^i o_t+v^i x_{t+1}+b^i)}{\partial b^i} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\widetilde{C}_{t+1}\sigma^{'}(u^i o_t+v^i x_{t+1}+b^i)\frac{\partial (u^i o_t+v^i x_{t+1}+b^i)}{\partial b^i} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\widetilde{C}_{t+1}\sigma^{'}(u^i o_t+v^i x_{t+1}+b^i)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Err}{\partial u^f} &= \sum_{t=0}^{T-1} \frac{\partial \frac{1}{2}(o_{t+1}-y_{t+1})^2}{\partial u^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial (o_{t+1}-y_{t+1})}{\partial u^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial o_{t+1}}{\partial u^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial (tanh(C_{t+1})h_{t+1})}{\partial u^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}\frac{\partial tanh(C_{t+1})}{\partial u^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\frac{\partial C_{t+1}}{\partial u^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\frac{\partial (f_{t+1}C_t+i_{t+1}\widetilde{C}_{t+1})}{\partial u^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})C_t\frac{\partial f_{t+1}}{\partial u^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})C_t\frac{\partial \sigma(u^f o_t+v^f x_{t+1}+b^f)}{\partial u^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})C_t\sigma^{'}(u^f o_t+v^f x_{t+1}+b^f)\frac{\partial (u^f o_t+v^f x_{t+1}+b^f)}{\partial u^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})C_t\sigma^{'}(u^f o_t+v^f x_{t+1}+b^f)o_t
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial Err}{\partial v^f} &= \sum_{t=0}^{T-1} \frac{\partial \frac{1}{2}(o_{t+1} - y_{t+1})^2}{\partial v^f} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial (o_{t+1} - y_{t+1})}{\partial v^f} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial o_{t+1}}{\partial v^f} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) \frac{\partial (tanh(C_{t+1}) h_{t+1})}{\partial v^f} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} \frac{\partial tanh(C_{t+1})}{\partial v^f} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \frac{\partial C_{t+1}}{\partial v^f} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) \frac{\partial (f_{t+1} C_t + i_{t+1} \widetilde{C}_{t+1})}{\partial v^f} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) C_t \frac{\partial f_{t+1}}{\partial v^f} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) C_t \frac{\partial \sigma(u^f o_t + v^f x_{t+1} + b^f)}{\partial v^f} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) C_t \sigma^{'}(u^f o_t + v^f x_{t+1} + b^f) \frac{\partial (u^f o_t + v^f x_{t+1} + b^f)}{\partial v^f} \\
&= \sum_{t=0}^{T-1} (o_{t+1} - y_{t+1}) h_{t+1} tanh^{'}(C_{t+1}) C_t \sigma^{'}(u^f o_t + v^f x_{t+1} + b^f) x_{t+1}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial Err}{\partial b^f} &= \sum_{t=0}^{T-1} \frac{\partial \frac{1}{2}(o_{t+1}-y_{t+1})^2}{\partial b^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial (o_{t+1}-y_{t+1})}{\partial b^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial o_{t+1}}{\partial b^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})\frac{\partial (tanh(C_{t+1})h_{t+1})}{\partial b^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}\frac{\partial tanh(C_{t+1})}{\partial b^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\frac{\partial C_{t+1}}{\partial b^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})\frac{\partial (f_{t+1}C_t + i_{t+1}\widetilde{C}_{t+1})}{\partial b^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})C_t\frac{\partial f_{t+1}}{\partial b^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})C_t\frac{\partial \sigma(u^f o_t + v^f x_{t+1} + b^f)}{\partial b^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})C_t\sigma^{'}(u^f o_t + v^f x_{t+1} + b^f)\frac{\partial (u^f o_t + v^f x_{t+1} + b^f)}{\partial b^f} \\
&= \sum_{t=0}^{T-1}(o_{t+1}-y_{t+1})h_{t+1}tanh^{'}(C_{t+1})C_t\sigma^{'}(u^f o_t + v^f x_{t+1} + b^f)
\end{aligned}
$$