确定系统参数

激励函数：sigmoid=1/(1+e^(-x)) 和 tanh=2Sigmoid(2x)-1=2/(1+e^(-2x))-1

学习率：实验确定，经验调参

隐层节点数：明文论文确定

训练次数：明文算法确定

1. 随机初始化权值

2. 信号前向传播 y=sigmoid(w.x)

沿着网络，从前向后依次计算输出层每个单元的输出

3. **误差反向传播△w=-αx(f(w.x)−y)**

（1）计算输出层误差

（2）计算隐藏层误差

4. **更新权值 w=w+△w**

5. 更新训练次数

6. 重复 2 到 5，直到迭代次数大于预先指定的学习次数

反向传播算法

基于误差的反向传播算法（backpropagation，BP）的前馈神经网络训练过程可以分为以下三步：

- 前馈计算每一层的净输入z^l 和激活值 a^l，直到最后一层
- 反向传播计算每一层的误差项
- 计算每一层参数的偏导数，并更新参数

其具体训练过程如下：

基于随机梯度下降的反向传播算法

输入: 训练集 $\mathcal{D}=\{(\mathbf{x}^{(n)},y^{(n)})\}_{n=1}^{N}$, 验证集 $\mathcal{V}$, 学习率 α, 正则化系数 λ, 网络层数 L, 神经元数量 $m^{(l)}$, $1\le l\le L$.

```
随机初始化 W, b ;
repeat
    对训练集 D 中的样本随机重排序;
    for n = 1 ··· N do
        从训练集 D 中选取样本 (x^(n), y^(n));
        前馈计算每一层的净输入 z^(l) 和激活值 a^(l)，直到最后一层;
        反向传播计算每一层的误差 δ^(l);            // 公式 (4.60)
        // 计算每一层参数的导数
        ∀l, ∂L(y^(n), ŷ^(n))/∂W^(l) = δ^(l)(a^(l-1))^T;   // 公式 (4.62)
        ∀l, ∂L(y^(n), ŷ^(n))/∂b^(l) = δ^(l);              // 公式 (4.63)
        // 更新参数
        W^(l) ← W^(l) − α(δ^(l)(a^(l-1))^T + λW^(l));
        b^(l) ← b^(l) − αδ^(l);
    end
until 神经网络模型在验证集 V 上的错误率不再下降;
```

输出: $W,\mathbf{b}$

主函数 Main.cpp
Network 类：{
网络初始化函数 init;
层添加函数 addLayer;
全网计算函数 compute;
网络前向计算函数 forwardNetwork;
网络反向计算函数 backwardNetwork;
Layer 类}
Layer 类：{
初始化函数 init，
激活函数 active，
梯度函数 gradient，
前向计算函数 forwardLayer
反向计算函数 backwardLayer}

1. 加载权重：loadWeight；将权重保存到每层：Network->layer->mWeight[][]
2. 保存权重：saveWeight
3. 训练网络：train；进行网络计算：network->compute
(1) 网络前向计算：forwardNetwork->层内前向计算：forwardLayer
(2) 计算误差 e，并返回当前网络权重参数 w
(3) 网络后向计算：backwardNetwork->层内反向计算：backwardLayer
(4) 权重参数更新：w=w+△w
4. 测试网络：test；进行网络计算：network->compute
(1) 网络前向计算：forwardNetwork->层内前向计算：forwardLayer
(2) 计算误差 e，并返回当前网络权重参数 w

电网数据集：分为训练集和测试集
- 明文实验

实验变量 1：逼近多项式次数 n（正确率，训练时间，训练次数）
实验变量 2：切比雪夫多项式 vs 勒让德多项式（正确率，训练时间，训练次数）
实验变量 3：sigmoid 函数 vs 逼近多项式函数（正确率，训练时间，训练次数）
选择上述逼近多项式中表现较好者作为隐私保护实验部分的备用
- 密文实验

1. 线性部分 u=**W.x**
安全乘法门：权重乘输入（调用三元组）
安全加法门：累加（本地）
2. 非线性部分 f(u) ->选择合适的逼近多项式替换
安全乘法门和安全加法门的组合

■ **LSTM (**x: scalar; **x**: vector; **X**: matrix**)**

#Plaintext algorithm

INPUT: $\mathbf{X}=\{\mathbf{x}_{(t-20)1}, \mathbf{x}_{(t-19)2}, \ldots, \mathbf{x}_{(t-1)20}\}$, where $\mathbf{x}_i \in R^{58}$ represents a 58-dimensional vector of real values at the *t* th time step, $\mathbf{x}_i=(e_i, \mathbf{i}, \mathbf{d}, \mathbf{h})=(e_i, i_1, ..., i_{48}, d_1, ..., d_7, h_0, h_1)$ [1+48+7+2=58]

OUTPUT: forecast $e'_{(t)21}$

#Initialization

0.Dataset

%Min-max normalisation for e_i, $e_{norm}=(e-e_{min})/(e_{max}-e_{min})$

Training set : Validating set : Testing set =7:2:1

1.Active function

Sigmoid: $\sigma=1/(1+e^{-x})$; $\sigma=a_3x^3+a_2x^2+a_1x+a_0$（To be determined）

dSigmoid: $\sigma'=\sigma(1-\sigma)$; $\sigma'=3a_3x^2+2a_2x+a_1$（To be determined）

Tanh: $\varphi=2/(1+e^{-2x})-1$; $\varphi=a_3x^3+a_2x^2+a_1x+a_0$ （To be determined）

dTanh: $\varphi'=1-\varphi^2$; $\varphi'=3a_3x^2+2a_2x+a_1$（To be determined）

ReLU: $\max\{x,0\}=(x>0)\cdot x$; $f=x^2$

dReLU: $f'(x)=(x>0)$; $f=2x$

2.Weight

%Layer1: LSTM (20 blocks)

$\mathbf{W}_{fx1\ 40X58}$; $\mathbf{W}_{fh1\ 40X40}$; $\mathbf{b}_{f1\ 40X1}$;

$\mathbf{W}_{ix1\ 40X58}$; $\mathbf{W}_{ih1\ 40X40}$; $\mathbf{b}_{i1\ 40X1}$;

$\mathbf{W}_{gx1\ 40X58}$; $\mathbf{W}_{gh1\ 40X40}$; $\mathbf{b}_{g1\ 40X1}$;

$\mathbf{W}_{ox1\ 40X58}$; $\mathbf{W}_{oh1\ 40X40}$; $\mathbf{b}_{o1\ 40X1}$;

%Layer2 : LSTM (20 blocks)

$\mathbf{W}_{fx2\ 40X40}$; $\mathbf{W}_{fh2\ 40X40}$; $\mathbf{b}_{f2\ 40X1}$;

$\mathbf{W}_{ix2\ 40X40}$; $\mathbf{W}_{ih2\ 40X40}$; $\mathbf{b}_{i2\ 40X1}$;

$\mathbf{W}_{gx2\ 40X40}$; $\mathbf{W}_{gh2\ 40X40}$; $\mathbf{b}_{g2\ 40X1}$;

$\mathbf{W}_{ox2\ 40X40}$; $\mathbf{W}_{oh2\ 40X40}$; $\mathbf{b}_{o2\ 40X1}$;

%Layer3: 20 hidden nodes

$\mathbf{W}_{3\ 20x40}$; $\mathbf{b}_{3\ 20x1}$

%Layer4: 1 output node

$\mathbf{W}_{4\ 1x20}$; $\mathbf{b}_{4\ 1x1}$

3.The memory cell state s_t and the intermediate output h_t

$\mathbf{h}_0=0$;

$\mathbf{s}_0=0$.

4.Hyper-parameters

Epochs=150

#Forward propagation

Layer1: LSTM (20 blocks)

For each block:

INPUT: $\mathbf{x}_{j\ 58X1}$ (j from 1 to 20)

$f_{j\ 40X1}=\sigma(\mathbf{W}_{fx1\ 40X58}.\mathbf{x}_{j\ 58X1}+\mathbf{W}_{fh1\ 40X40}.\mathbf{h}_{j-1\ 40X1}+\mathbf{b}_{f1\ 40X1})$

$\mathbf{i}_{j\ 40X1}=\sigma(\mathbf{W}_{ix1\ 40X58}.\mathbf{x}_{j\ 58X1}+\mathbf{W}_{ih1\ 40X40}.\mathbf{h}_{j-1\ 40X1}+\mathbf{b}_{i1\ 40X1})$

$\mathbf{g}_{j\ 40X1}=\varphi(\mathbf{W}_{gx1\ 40X58}.\mathbf{x}_{j\ 58X1}+\mathbf{W}_{gh1\ 40X40}.\mathbf{h}_{j-1\ 40X1}+\mathbf{b}_{g1\ 40X1})$

$\mathbf{o}_{j\ 40X1}=\sigma(\mathbf{W}_{ox1\ 40X58}.\mathbf{x}_{j\ 58X1}+\mathbf{W}_{oh1\ 40X40}.\mathbf{h}_{j-1\ 40X1}+\mathbf{b}_{o1\ 40X1})$

$\mathbf{s}_{j\ 40X1} = \mathbf{g}_{j\ 40X1}\odot\mathbf{i}_{j\ 40X1}+ \mathbf{s}_{j-1\ 40X1}\odot\mathbf{f}_{j\ 40X1}$

$\mathbf{h}_{j\ 40X1} = \mathbf{s}_{j\ 40X1}\odot\mathbf{o}_{j\ 40X1}$

OUTPUT: $\mathbf{h}_{j\ 40X1}$

For Layer1: $\mathbf{h}_{1\ 40X1}, \ldots, \mathbf{h}_{20\ 40X1}$

Layer2: LSTM (20 blocks)

For each block:

INPUT: $\mathbf{h}_{j\ 40X1}$(j from 1 to 20)

$f_{j\ 40X1}=\sigma(\mathbf{W}_{fx2\ 40X40}.\mathbf{h}_{j\ 40X1}+\mathbf{W}_{fh2\ 40X40}.\mathbf{h}_{j-1\ 40X1}+\mathbf{b}_{f2\ 40X1})$

$\mathbf{i}_{j\ 40X1}=\sigma(\mathbf{W}_{ix2\ 40X40}.\mathbf{h}_{j\ 40X1}+\mathbf{W}_{ih2\ 40X40}.\mathbf{h}_{j-1\ 40X1}+\mathbf{b}_{i2\ 40X1})$

$\mathbf{g}_{j\ 40X1}=\varphi(\mathbf{W}_{gx2\ 40X40}.\mathbf{h}_{j\ 40X1}+\mathbf{W}_{gh2\ 40X40}.\mathbf{h}_{j-1\ 40X1}+\mathbf{b}_{g2\ 40X1})$

$\mathbf{o}_{j\ 40X1}=\sigma(\mathbf{W}_{ox2\ 40X40}.\mathbf{h}_{j\ 40X1}+\mathbf{W}_{oh2\ 40X40}.\mathbf{h}_{j-1\ 40X1}+\mathbf{b}_{o2\ 40X1})$

$\mathbf{s}_{j40X1} = \mathbf{g}_{j\ 40X1}\odot\mathbf{i}_{j\ 40X1}+ \mathbf{s}_{j-1\ 40X1}\odot\mathbf{f}_{j\ 40X1}$

$\mathbf{h}_{j\ 40X1} = \mathbf{s}_{j\ 40X1}\odot\mathbf{o}_{j\ 40X1}$

OUTPUT: $\mathbf{h}_{j\ 40X1}$

For Layer2: $\mathbf{h}_{20\ 40X1}$

$\mathbf{Z}_{240x1}=\mathbf{h}_{20\ 40X1}$

$\mathbf{A}_{240x1}=\mathbf{Z}_{20\ 40X1}$

Layer3: Dense (20 hidden nodes)

INPUT: $\mathbf{h}_{20\ 40X1}$

$\mathbf{Z}_{3\ 20x1}=\mathbf{W}_{3\ 20x40}.\mathbf{h}_{20\ 40X1}+\mathbf{b}_{3\ 20x1}$

$\mathbf{A}_{3\ 20x1}=$ RELU$(\mathbf{Z}_{3\ 20x1})$

OUTPUT: $\mathbf{A}_{3\ 20x1}$

Layer4: Dense (a single value: forecasting)

INPUT: $\mathbf{A}_{3\ 20x1}$

$\mathbf{Z}_{4\ 1x1}=\mathbf{W}_{4\ 1x20}.\mathbf{A}_{3\ 20x1}+\mathbf{b}_{4\ 1x1}$

$e'_{21\ 1x1}=$ RELU($\mathbf{Z}_{4\ 1x1}$)

OUTPUT: $e'_{21\ 1x1}$

%we can replace RELU with Sigmoid, Tanh, or even Swish=x.Sigmoid

#Backpropagation:SGD

%To learn the coefficients w, a cost function C(w) is defined and w is calculated by the optimization $\text{argmin}_w C(w)$. In LSTM, a commonly used cost function is MSE.

%MSE: $C(w)=(1/n).\text{sum}\{i=1:n\}C_i(w)$; where $C_i(w)=1/2(e'_t - e_t)^2$.

% A sample (x_i, y_i) is selected randomly and a coefficient w_j is updated as

%$w_j = w_j - \alpha(\partial C_i(w)/\partial wj)$

Layer4: Dense (a single value: forecasting)

$G_{41x1} = \partial A_4/\partial Z_4 = 1-\varphi^2(Z_{4\ 1x1})$

$\partial C(w)/\partial A_{4\ 1x1} = e'_t - e_t$

$\partial C(w)/\partial Z_{4\ 1x1} = (\partial C(w)/\partial A_4)(\partial A_4/\partial Z_4) = (e'_t - e_t)(1-\varphi^2(Z_4))$

$\triangle\mathbf{W}_{41x20} = \partial C(w)/\partial W_4 = (\partial C(w)/\partial Z_4)_{1x1}\ A_3{}^T{}_{1x20}$

$\triangle\mathbf{b}_{4\ 1x1} = \partial C(w)/\partial b_4 = (\partial C(w)/\partial Z_4)_{1x1}$

Update: $\mathbf{W}_{4\ 1x20}=\mathbf{W}_{4\ 1x20}-\alpha\triangle\mathbf{W}_{41x20}$; $\mathbf{b}_{4\ 1x1}=\mathbf{b}_{4\ 1x1}-\alpha\triangle\mathbf{b}_{4\ 1x1}$, where α is the learning rat.

Layer3: Dense (20 hidden nodes)

$G_{3\,20x20}=\partial A_3/\partial Z_3=\{1-\varphi^2(Z_{3\,20x1})$: Only the diagonal of the square matrix has a value$\}$

$\partial C(w)/\partial Z_{3\,20x1}=G_{3\,20x20}W_4{}^{T}{}_{20x1}\,(\partial C(w)/\partial Z_4)_{1x1}$

$\triangle\mathbf{W}_{3\,20x40}=\partial C(w)/\partial W_3=(\partial C(w)/\partial Z_3)_{20x1}\,\mathbf{A}_2{}^{T}{}_{1x40}$

$\triangle\mathbf{b}_{3\,20x1}=\partial C(w)/\partial b_3=(\partial C(w)/\partial Z_3)_{20x1}$

Update: $\mathbf{W}_{3\,20x40}=\mathbf{W}_{3\,20x40}-\alpha\triangle\mathbf{W}_{3\,20x40}$

Update: $\mathbf{b}_{3\,20x1}=\mathbf{b}_{3\,20x1}-\alpha\triangle\mathbf{b}_{3\,20x1}$

Layer2: LSTM (20 blocks)

$G_{2\,40x40}=\partial A_2/\partial Z_2=I_{40}$

$\partial C(w)/\partial h_{20\,40x1}=\partial C(w)/\partial Z_{2\,40x1}=G_{2\,40x40}W_3{}^{T}{}_{40x20}(\partial C(w)/\partial Z_3)_{20x1}$

$\partial C(w)/\partial s_{20\,40x1}=(\partial C(w)/\partial h_{20})_{40x1}\odot o_{20\,40X1}$

For t from 20 to 1 {

$\partial C(w)/\partial f_{t\,40x1}=(\partial C(w)/\partial s_t)_{40x1}\odot s_{t-1\,40x1}$

$\partial C(w)/\partial i_{t\,40x1}=(\partial C(w)/\partial s_t)_{40x1}\odot g_{t\,40x1}$

$\partial C(w)/\partial g_{t\,40x1}=(\partial C(w)/\partial s_t)_{40x1}\odot i_{t\,40x1}$

$\partial C(w)/\partial o_{t\,40x1}=(\partial C(w)/\partial h_t)_{40x1}\odot s_{t\,40x1}$

$\partial C(w)/\partial f'_{t\,40x1}=(\partial C(w)/\partial f_t)_{40x1}.f_t(1-f_t)$

$\partial C(w)/\partial i'_{t\,40x1}=(\partial C(w)/\partial i_t)_{40x1}.i_t(1-i_t)$

$\partial C(w)/\partial g'_{t\,40x1}=(\partial C(w)/\partial g_t)_{40x1}.(1-g_t{}^2)$

$\partial C(w)/\partial o'_{t\,40x1}=(\partial C(w)/\partial o_t)_{40x1}.o_t(1-o_t)$

$\partial C(w)/\partial x_{20\,40x1}=\mathbf{W}_{fx2}{}^{T}{}_{40x40}.(\partial C(w)/\partial f'_t)_{40x1}+\mathbf{W}_{ix2}{}^{T}{}_{40x40}.(\partial C(w)/\partial i'_t)_{40x1}+$
$\mathbf{W}_{gx2}{}^{T}{}_{40x40}.(\partial C(w)/\partial g'_t)_{40x1}+\mathbf{W}_{ox2}{}^{T}{}_{40x40}.(\partial C(w)/\partial o'_t)_{40x1}$

$\partial C(w)/\partial\mathbf{W}_{fx2t}=(\partial C(w)/\partial f'_t)_{40x1}.\mathbf{x}_t{}^{T}{}_{1X40}$

$\partial C(w)/\partial\mathbf{W}_{fh2t}=(\partial C(w)/\partial f'_t)_{40x1}.\mathbf{h}_{t-1}{}^{T}{}_{1X40}$

$\partial C(w)/\partial\mathbf{W}_{ix2t}=(\partial C(w)/\partial i'_t).\mathbf{x}_t{}^{T}{}_{1X40}$

$\partial C(w)/\partial\mathbf{W}_{ih2t}=(\partial C(w)/\partial i'_t).\mathbf{h}_{t-1}{}^{T}{}_{1X40}$

$\partial C(w)/\partial\mathbf{W}_{gx2t}=(\partial C(w)/\partial g'_t).\mathbf{x}_t{}^{T}{}_{1X40}$

$\partial C(w)/\partial\mathbf{W}_{gh2t}=(\partial C(w)/\partial g'_t).\mathbf{h}_{t-1}{}^{T}{}_{1X40}$

$\partial C(w)/\partial\mathbf{W}_{ox2t}=(\partial C(w)/\partial o'_t).\mathbf{x}_t{}^{T}{}_{1X40}$

$\partial C(w)/\partial\mathbf{W}_{oh2t}=(\partial C(w)/\partial o'_t).\mathbf{h}_{t-1}{}^{T}{}_{1X40}$

$t=t-1$

$\partial C(w)/\partial h_{t\,40x1}=\mathbf{W}_{fh2}{}^{T}{}_{40x40}.(\partial C(w)/\partial f'_{t+1})_{40x1}+\mathbf{W}_{ih2}{}^{T}{}_{40x40}.(\partial C(w)/\partial i'_{t+1})_{40x1}+$
$\mathbf{W}_{gh2}{}^{T}{}_{40x40}.(\partial C(w)/\partial g'_{t+1})_{40x1}+\mathbf{W}_{oh2}{}^{T}{}_{40x40}.(\partial C(w)/\partial o'_{t+1})_{40x1}$

$\partial C(w)/\partial s_{t\,40x1}=(\partial C(w)/\partial h_t)_{40x1}\odot o_{t\,40x1}+(\partial C(w)/\partial s_{t+1})_{40x1}\odot f_{t+1\,40x1}$

}

$\triangle\mathbf{W}_{fx2}=\partial C(w)/\partial\mathbf{W}_{fx2}=\mathrm{sum}\{\partial C(w)/\partial\mathbf{W}_{fx2t}|t=1:20\};$

$\triangle\mathbf{W}_{fh2}=\partial C(w)/\partial\mathbf{W}_{fh2}=\mathrm{sum}\{\partial C(w)/\partial\mathbf{W}_{fh2t}|t=1:20\};$

$\triangle\mathbf{W}_{ix2}=\partial C(w)/\partial\mathbf{W}_{ix2}=\mathrm{sum}\{\partial C(w)/\partial\mathbf{W}_{ix2t}|t=1:20\};$

$\triangle\mathbf{W}_{ih2}=\partial C(w)/\partial\mathbf{W}_{ih2}=\mathrm{sum}\{\partial C(w)/\partial\mathbf{W}_{ih2t}|t=1:20\};$

$\triangle\mathbf{W}_{gx2}=\partial C(w)/\partial\mathbf{W}_{gx2}=\mathrm{sum}\{\partial C(w)/\partial\mathbf{W}_{gx2t}|t=1:20\};$

$\triangle\mathbf{W}_{gh2}=\partial C(w)/\partial\mathbf{W}_{gh2}=\mathrm{sum}\{\partial C(w)/\partial\mathbf{W}_{gh2t}|t=1:20\};$

$\triangle \mathbf{W}_{ox2}=\partial C(w)/\partial \mathbf{W}_{ox2}=\text{sum}\{\partial C(w)/\partial \mathbf{W}_{ox2t}|t=1:20\}$;

$\triangle \mathbf{W}_{oh2}=\partial C(w)/\partial \mathbf{W}_{oh2}=\text{sum}\{\partial C(w)/\partial \mathbf{W}_{oh2t}|t=1:20\}$.

Update

Layer1: LSTM (20 blocks)

$\partial C(w)/\partial h_{20\,40x1}=\partial C(w)/\partial x_{20\,40x1}$

$\partial C(w)/\partial s_{20\,40x1}=(\partial C(w)/\partial h_t)_{40x1}\odot o_{t\,40x1}$

For t from 20 to 1 {

$\partial C(w)/\partial f_{t\,40x1}=(\partial C(w)/\partial s_t)_{40x1}\odot s_{t-1\,40x1}$

$\partial C(w)/\partial i_{t\,40x1}=(\partial C(w)/\partial s_t)_{40x1}\odot g_{t\,40x1}$

$\partial C(w)/\partial g_{t\,40x1}=(\partial C(w)/\partial s_t)_{40x1}\odot i_{t\,40x1}$

$\partial C(w)/\partial o_{t\,40x1}=(\partial C(w)/\partial h_t)_{40x1}\odot s_{t\,40x1}$

$\partial C(w)/\partial f'_{t\,40x1}=(\partial C(w)/\partial f_t)_{40x1}.f_t(1-f_t)$

$\partial C(w)/\partial i'_{t\,40x1}=(\partial C(w)/\partial i_t)_{40x1}.i_t(1-i_t)$

$\partial C(w)/\partial g'_{t\,40x1}=(\partial C(w)/\partial g_t)_{40x1}.(1-g_t^2)$

$\partial C(w)/\partial o'_{t\,40x1}=(\partial C(w)/\partial o_t)_{40x1}.o_t(1-o_t)$

$\partial C(w)/\partial \mathbf{W}_{fx1t}=(\partial C(w)/\partial f'_t)_{40x1}.\mathbf{x}_t{}^{T}{}_{1X40}$

$\partial C(w)/\partial \mathbf{W}_{fh1t}=(\partial C(w)/\partial f'_t)_{40x1}.\mathbf{h}_{t-1}{}^{T}{}_{1X40}$

$\partial C(w)/\partial \mathbf{W}_{ix1t}=(\partial C(w)/\partial i'_t).\mathbf{x}_t{}^{T}{}_{1X40}$

$\partial C(w)/\partial \mathbf{W}_{ih1t}=(\partial C(w)/\partial i'_t).\mathbf{h}_{t-1}{}^{T}{}_{1X40}$

$\partial C(w)/\partial \mathbf{W}_{gx1t}=(\partial C(w)/\partial g'_t).\mathbf{x}_t{}^{T}{}_{1X40}$

$\partial C(w)/\partial \mathbf{W}_{gh1t}=(\partial C(w)/\partial g'_t).\mathbf{h}_{t-1}{}^{T}{}_{1X40}$

$\partial C(w)/\partial \mathbf{W}_{ox1t}=(\partial C(w)/\partial o'_t).\mathbf{x}_t{}^{T}{}_{1X40}$

$\partial C(w)/\partial \mathbf{W}_{oh1t}=(\partial C(w)/\partial o'_t).\mathbf{h}_{t-1}{}^{T}{}_{1X40}$

t=t-1

$\partial C(w)/\partial h_{t\,40x1}=\mathbf{W}_{fx2}{}^{T}{}_{40x40}.(\partial C(w)/\partial f'_t)_{40x1}+\mathbf{W}_{ix2}{}^{T}{}_{40x40}.(\partial C(w)/\partial i'_t)_{40x1}+$
$\mathbf{W}_{gx2}{}^{T}{}_{40x40}.(\partial C(w)/\partial g'_t)_{40x1}+\mathbf{W}_{ox2}{}^{T}{}_{40x40}.(\partial C(w)/\partial o'_t)_{40x1}+$
$\mathbf{W}_{fh1}{}^{T}{}_{40x40}.(\partial C(w)/\partial f'_{t+1})_{40x1}+\mathbf{W}_{ih1}{}^{T}{}_{40x40}.(\partial C(w)/\partial i'_{t+1})_{40x1}+$
$\mathbf{W}_{gh1}{}^{T}{}_{40x40}.(\partial C(w)/\partial g'_{t+1})_{40x1}+\mathbf{W}_{oh1}{}^{T}{}_{40x40}.(\partial C(w)/\partial o'_{t+1})_{40x1}$

$\partial C(w)/\partial s_{t\,40x1}=(\partial C(w)/\partial h_t)_{40x1}\odot o_{t\,40x1}+(\partial C(w)/\partial s_{t+1})_{40x1}\odot f_{t+1\,40x1}$

}

$\triangle \mathbf{W}_{fx1}=\partial C(w)/\partial \mathbf{W}_{fx1}=\text{sum}\{\partial C(w)/\partial \mathbf{W}_{fx1t}|t=1:20\}$;

$\triangle \mathbf{W}_{fh1}=\partial C(w)/\partial \mathbf{W}_{fh1}=\text{sum}\{\partial C(w)/\partial \mathbf{W}_{fh1t}|t=1:20\}$;

$\triangle \mathbf{W}_{ix1}=\partial C(w)/\partial \mathbf{W}_{ix1}=\text{sum}\{\partial C(w)/\partial \mathbf{W}_{ix1t}|t=1:20\}$;

$\triangle \mathbf{W}_{ih1}=\partial C(w)/\partial \mathbf{W}_{ih1}=\text{sum}\{\partial C(w)/\partial \mathbf{W}_{ih1t}|t=1:20\}$;

$\triangle \mathbf{W}_{gx1}=\partial C(w)/\partial \mathbf{W}_{gx1}=\text{sum}\{\partial C(w)/\partial \mathbf{W}_{gx1t}|t=1:20\}$;

$\triangle \mathbf{W}_{gh1}=\partial C(w)/\partial \mathbf{W}_{gh1}=\text{sum}\{\partial C(w)/\partial \mathbf{W}_{gh1t}|t=1:20\}$;

$\triangle \mathbf{W}_{ox1}=\partial C(w)/\partial \mathbf{W}_{ox1}=\text{sum}\{\partial C(w)/\partial \mathbf{W}_{ox1t}|t=1:20\}$;

$\triangle \mathbf{W}_{oh1}=\partial C(w)/\partial \mathbf{W}_{oh1}=\text{sum}\{\partial C(w)/\partial \mathbf{W}_{oh1t}|t=1:20\}$.

Update