1.

(a).

$$L = \sum_t \log P(y_t | \vec{x_t})$$

$$= \sum_t \log \left[p_t^{\,y_t} (1-p_t)^{1-y_t} \right]$$

$$= \sum_t \left[y_t \log p_t + (1-y_t) \log (1-p_t) \right]$$

$$\frac{\partial L}{\partial w_i} = \sum_t \left[\frac{y_t}{p_t} \frac{\partial p_t}{\partial w_i} + \frac{(1-y_t)}{(1-p_t)} \frac{\partial (1-p_t)}{\partial w_i} \right]$$

$$= \sum_t \left[\frac{y_t}{p_t} \frac{\partial p_t}{\partial w_i} - \frac{(1-y_t)}{(1-p_t)} \frac{\partial p_t}{\partial w_i} \right]$$

$$= \sum_t \left[\left(\frac{y_t}{p_t} - \frac{1-y_t}{1-p_t} \right) \cdot \frac{\partial p_t}{\partial w_i} \right]$$

$$\because \frac{\partial p_t}{\partial w_i} = \frac{\partial g(\vec{w}, \vec{x})}{\partial w_i, \partial x_i} \cdot \partial x_i = g'(\vec{w} \cdot \vec{x_t}) \cdot x_i.$$

$$\because \frac{y_t}{p_t} - \frac{1-y_t}{1-p_t} = \frac{y_t(1-p_t) - (1-y_t)p_t}{p_t(1-p_t)}$$

$$= \frac{y_t - y_t p_t - p_t + y_t p_t}{p_t(1-p_t)} = \frac{y_t - p_t}{p_t(1-p_t)}$$

$$\therefore \frac{\partial L}{\partial w_i} = \sum_t \frac{y_t - p_t}{p_t(1-p_t)} \cdot g'(\vec{w} \cdot \vec{x_i}) \cdot x_{it}.$$

(b). $\because g(z) = [1+e^{-z}]^{-1}$

$\therefore g(z)' = g(z)g(-z)$; $g(-z) = 1-g(z)$

$\because p_t = P(Y=1|\vec{x_t}) = g(\vec{w}\cdot\vec{x})$

$1-p_t = g(-\vec{w}\vec{x})$.

$\therefore g(\vec{w}\vec{x})' = p_t(1-p_t)$.

$\therefore$

$$\frac{\partial L}{\partial w_i} = \sum_{t=1}^{T}\left[\frac{p_t(1-p_t)}{p_t(1-p_t)}\right]\cdot(y_t-p_t)x_{it}.$$

$$= \sum_{t=1}^{T}(y_t - g(\vec{w}\cdot\vec{x}))\cdot\vec{x_t}$$

2.

$L = \sum_t \log P(y_t|\vec{x_t})$.

as for $P(Y=i|\vec{x}) = \dfrac{e^{\vec{w_i}\vec{x_i}}}{\sum_j e^{\vec{w_j}\vec{x}}}$.

and there are C kinds of Y.

$$\therefore L = \sum_t \log\left(\frac{\prod_{i=1}^{C}(e^{\vec{w_i}\vec{x}})^{y_{it}}}{\sum e^{\vec{w_j}\vec{x}}}\right)$$

$$= \sum_t\left(\sum_{i=1}^{C}\vec{w_i}\vec{x}\cdot y_{it} - \log\sum e^{\vec{w_j}\vec{x}}\right)$$

$$\therefore \frac{\partial L}{\partial \vec{w}} = \sum\left[y_{it}\vec{x} - \frac{1}{\sum_{j=1} e^{\vec{w_j}\vec{x}}}\cdot\vec{x}\cdot e^{\vec{w_j}\vec{x}}\right]$$

$$= \sum_t\left[(y_{it} - p_{it})\cdot\vec{x}\right]$$

3.

(a) $\varepsilon_n = x_n - x_*$

$f'(x) = \alpha(x - x_*) = \alpha\varepsilon_n$

$\because$ $\varepsilon_n = x_n - x_*$, $x_n = x_{n-1} - \eta \cdot \alpha\varepsilon_{n-1}$

$\therefore$ $\varepsilon_n = x_{n-1} - \eta\alpha\varepsilon_{n-1} - x_*$

$= (x_{n-1} - x_*) - \eta\alpha\varepsilon_{n-1}$

$= \varepsilon_{n-1} - n\alpha\varepsilon_{n-1} = (1-n\alpha)\varepsilon_{n-1}$

$= (1-n\alpha)[(1-n\alpha)\varepsilon_{n-2}] \cdots$

$= (1-n\alpha)^n \varepsilon_0$

(b) we can see from part (a). $\varepsilon_n \; \alpha \; \varepsilon_0$,

$\therefore$ if $|(1-n\alpha)| < 0$, then it converges.

as for $f''(x)$:

$f''(x) = \alpha$.

when $1-n\alpha = 0 \Rightarrow n = \frac{1}{\alpha}$, it converge fast.

$\therefore$ when $f''(x) = \alpha$ and $n = \frac{1}{\alpha}$, converge fast

(c) $\varepsilon_{n+1} = x_{n+1} - x_*$

$= x_n - \eta f'(x_n) + \beta(x_n - x_{n-1}) - x_*$

$= (1-\eta\alpha)(x_n - x_*) + \beta(x_n - x_{n-1})$

$= (1-\eta\alpha)\varepsilon_n + \beta((x_n - x_*) - (x_{n-1} - x_*))$

$= (1-\eta\alpha)\varepsilon_n + \beta(\varepsilon_n - \varepsilon_{n-1})$

$= (1-\eta\alpha+\beta)\varepsilon_n - \beta\varepsilon_{n-1}$

1d). $\alpha=1,\ \eta=\frac{4}{9},\ \beta=\frac{1}{9}$

$$\varepsilon_{n+1}=(1-\alpha\eta+\beta)\varepsilon_n-\beta\varepsilon_{n-1}$$
$$=\frac{6}{9}\varepsilon_n-\frac{1}{9}\varepsilon_{n-1}$$
$$=\frac{2}{3}\varepsilon_n-\frac{1}{9}\varepsilon_{n-1}$$

assume $\varepsilon_n=\lambda^n\varepsilon_0$.

$\therefore$

$$\lambda^{n+1}\varepsilon_0=\frac{2}{3}\lambda^n\varepsilon_0-\frac{1}{9}\lambda^{n-1}\varepsilon_0$$
$$\therefore\ 9\lambda^2-6\lambda+1=0$$
$$\lambda=\frac{1}{3}.$$
$$\therefore\ \varepsilon_n=\frac{1}{3}^n\varepsilon_0$$

And:

as for $\eta=\frac{4}{9},\ \beta=0,\ \varepsilon_n=(\frac{5}{9})^n\varepsilon_0$, which faster.

4

(a). $$\varepsilon_n=|x_n-x_*|$$
$$=\left|x_{n-1}-\frac{g'(x_{n-1})}{g''(x_{n-1})}-x_*\right|$$
$$=\left|x_{n-1}-\frac{\cancel{2k}(x_{n-1}-x_*)^{\cancel{2k-1}}}{\cancel{2k}\cdot(2k-1)\cdot\cancel{(x_{n-1}-x_*)^{2k-2}}}-x_*\right|$$
$$=\left|(1-\frac{1}{2k-1})(x_{n-1}-x_*)\right|$$
$$=(1-\frac{1}{2k-1})\cdot\varepsilon_{n-1}=(1-\frac{1}{2k-1})^n\varepsilon_0.$$

(b). $\because \varepsilon_n \le \delta \varepsilon_0$.

$\therefore \dfrac{\varepsilon_n}{\varepsilon_0} \le \delta$

$\therefore \left(\dfrac{2k-2}{2k-1}\right)^n \le \delta$

$\therefore n \log\left(\dfrac{2k-2}{2k-1}\right) \le \log(\delta)$

$\therefore n \left(\dfrac{2k-2}{2k-1} - 1\right) \le \log(\delta)$

$\therefore n \left(\dfrac{-1}{2k-1}\right) \le \log(\delta)$

$\therefore n \ge -(2k-1)\log(\delta) = (2k-1)\log\left(\dfrac{1}{\delta}\right)$

(c) $h(x) = x_* \log(x_*/x) - x_* + x$

$$h'(x) = x_* \frac{d}{dx}\left(\log \frac{x_*}{x}\right) + 1$$

$$= x_* \cdot \frac{\cancel{x}}{\cancel{x_*}} \cdot \cancel{x_*} \cdot \frac{1}{-x^2} + 1$$

$$= 1 - \frac{x_*}{x}$$

$$h''(x) = -x_* \cdot -\frac{1}{x^2} = \frac{x_*}{x^2}$$

$\therefore$ when $h'(x) = 0$, $h''(x) > 0$, get the minimum:

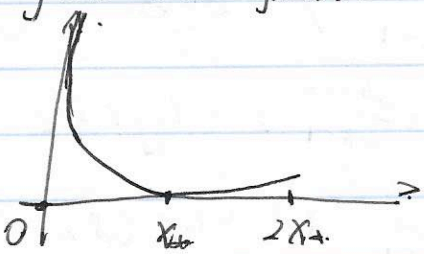
$1 - \dfrac{x_*}{x} = 0$, $\quad x = x_*$

graph:

since x_* is any number > 0.

after pick up several number $x_* = 5, 100, 1000$,

and get the function plots should like.



(d). $x_{n+1} = x_n - \frac{g'(x_n)}{g''(x_n)}$

$$= x_n - \frac{1 - \frac{x_*}{x}}{\frac{x_*}{x^2}} = x_n - \frac{x - x_*}{x} \cdot \frac{x^2}{x_*}$$

$$= x_n - \frac{x_n(x_n - x_*)}{x_*} = \frac{x_n x_* - x_n^2 + x_n x_*}{x_*}$$

$$x_{n+1} = \frac{-x_n^2 + 2x_n x_* - x_*^2}{x_*} + \frac{x_*^2}{x_*}$$

$$\therefore x_{n+1} - x_* = -\frac{(x_n - x_*)^2}{x_*}$$

$$\therefore \frac{x_{n+1} - x_*}{x_*} = e_{n+1} = -\left(\frac{x_n - x_*}{x_*}\right)^2 = -e_n^2$$

$$\therefore e_{n+1} = -e_n^2 = -(e_0)^{2n}$$

$$\therefore |e_0| < 1 \Rightarrow 0 < x < 2x_*$$