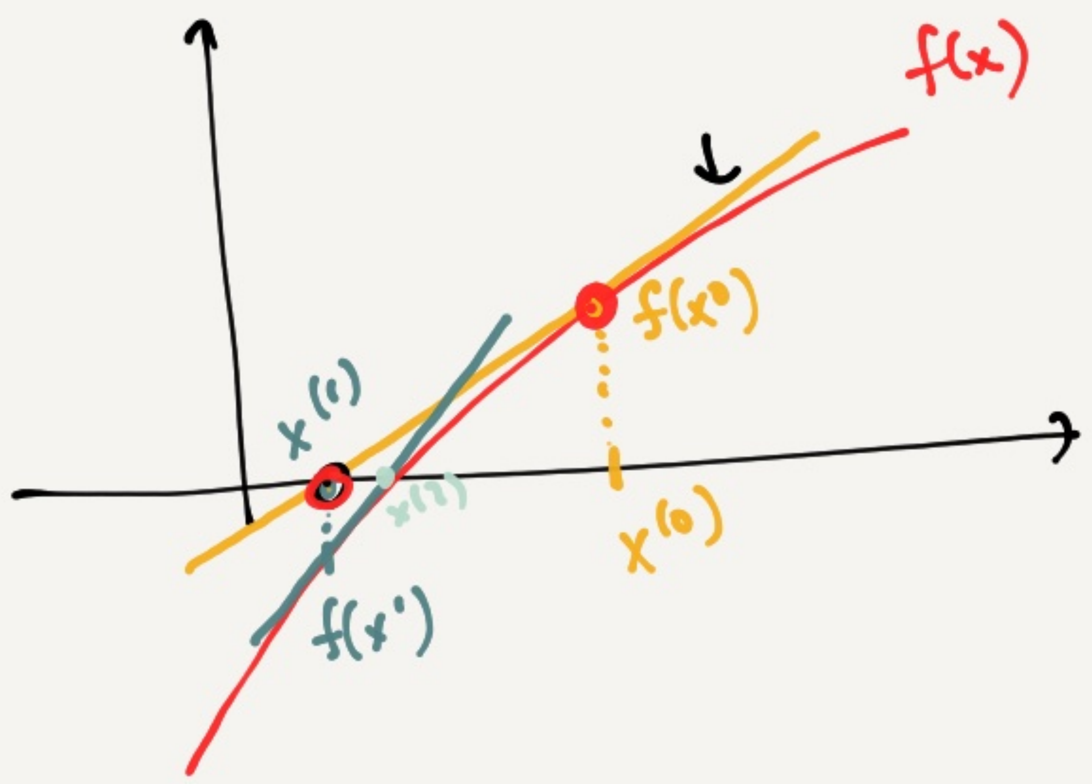
Tutorial 8

Problem 8.6 - Newton-Raphson Method

Goal: find the zero crossing of $f(x)$,
i.e. $f(x^*)=0$



Given $x^{(0)}$, find the zero-crossing of the line tangent to $f(x^{(0)})$:

$$f'(x^{(0)}) = \frac{\Delta f}{\Delta x} = \frac{f(x^{(0)}) - 0}{x^{(0)} - \underbrace{x^{(1)}}}$$

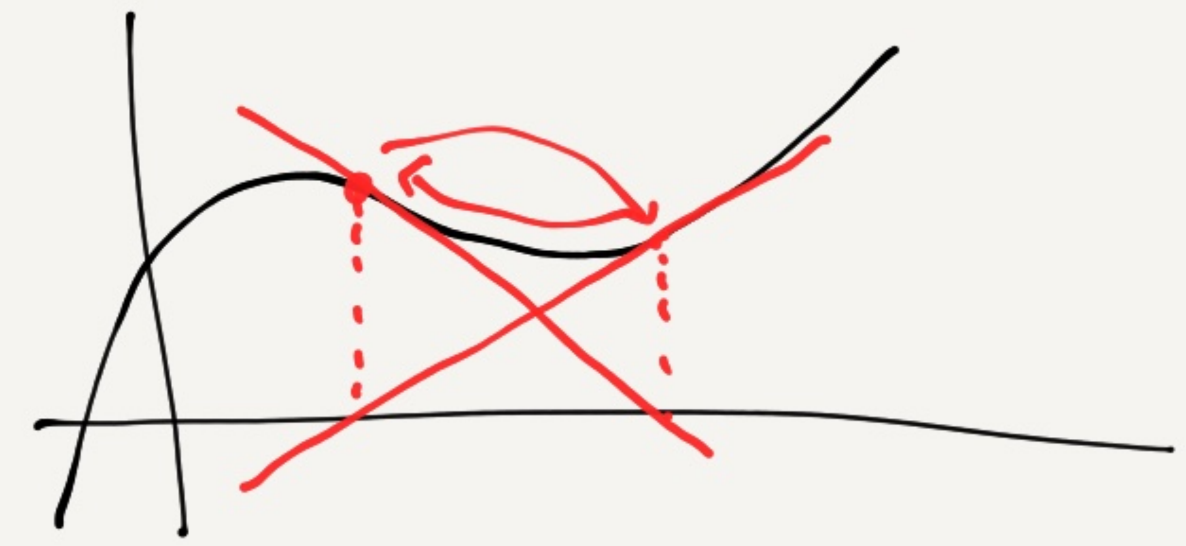
$$f'(x^{(0)}) = \frac{f(x^{(0)})}{x^{(0)} - x^{(1)}}$$

$$x^{(0)} - x^{(1)} = \frac{f(x^{(0)})}{f'(x^{(0)})} \Rightarrow \boxed{x^{(1)} = x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})}}$$

- Repeat until convergence $\Rightarrow x^*$ for $f(x^*)=0$
- For well-behaved functions it converges if you start sufficiently close to the x^*.

(sometimes doesn't converge (part b))

e.g.



c) How to apply this to minimize $g(x)$?

let $f(x) = g'(x)$, find the zero-crossing of $f(x)$.

$$\Rightarrow \boxed{x^{(i+1)} = x^{(i)} - \frac{g'(x^{(i)})}{g''(x^{(i)})}}$$

like GD except learning rate is the 2nd derivative

CS5487 Lecture Notes (2020)
Dr. Antoni B. Chan
Dept of Computer Science
City University of Hong Kong

Problem 8-4

Regularized LR:

likelihood: $p(y_i|x_i,w) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$

$$\pi_i = \sigma(w^T x_i)$$

prior $p(w) = N(w|0,\Gamma^{-1})$

Find MAP solution:

$$w^* = \arg\max_w \log p(y|X,w) + \log p(w)$$

$$= \arg\max_w \left[\sum_{i=1}^N y_i \log \pi_i + (1-y_i)\log(1-\pi_i)\right] - \frac{1}{2} w^T\Gamma w + \dots$$

$$= \arg\min_w \underbrace{\underbrace{\left[-\sum_i y_i\log\pi_i + (1-y_i)\log(1-\pi_i)\right]}_{\hat{E}(w)} + \frac{1}{2} w^T\Gamma w}_{E(w)}$$

Find the gradient & Hessian of $E(w)$:

$$\frac{\partial}{\partial w}\hat{E}(w) = \frac{\partial}{\partial w} -\sum_i y_i \log \pi_i + (1-y_i)\log(1-\pi_i)$$

$$= -\sum_i y_i \frac{1}{\pi_i}\frac{\partial \pi_i}{\partial w} + (1-y_i)\frac{1}{(1-\pi_i)}\left(-\frac{\partial\pi_i}{\partial w}\right)$$

$$\frac{\partial \pi_i}{\partial w} = \frac{\partial}{\partial w}\sigma(w^Tx_i) = \underbrace{\sigma'(w^Tx_i)}_{\text{PS 8-1}} \frac{\partial w^T x_i}{\partial w}$$

PS 8-1
$\sigma'(x) = \sigma(x)(1-\sigma(x))$

$$= \underbrace{\sigma(w^Tx_i)}_{\pi_i}\underbrace{(1-\sigma(w^Tx_i))}_{1-\pi_i}\cdot x_i = \pi_i(1-\pi_i)x_i$$

$$\Rightarrow \quad = -\sum_i y_i(1-\pi_i)x_i + (1-y_i)(-\pi_i x_i)$$

$$= -\sum_i \left(y_i x_i - \cancel{y_i\pi_i x_i} - \pi_i x_i + \cancel{y_i\pi_i x_i}\right)$$

$$= \sum_i \pi_i x_i - y_i x_i$$

$$= \sum_i(\pi_i - y_i)x_i = \boxed{X(\pi - y)}$$

$X = [x_1 \cdots x_N]$

$\pi = \begin{bmatrix}\pi_1\\ \vdots \\ \pi_N\end{bmatrix}$

$y = \begin{bmatrix}y_1\\ \vdots \\ y_N\end{bmatrix}$

$$\Rightarrow \boxed{\nabla E(w) = X(\pi-y) + \Gamma w}$$

Hessian of $E(w)$

$$\nabla^2 E(w) = \frac{\partial}{\partial w}\left[\frac{\partial}{\partial w^T}E(w)\right]$$

$$= \frac{\partial}{\partial w}\left[(\pi-y)^TX^T + w^T\Gamma\right]$$

$$= \frac{\partial}{\partial w}\left[\pi^TX^T - \cancel{y^TX^T} + w^T\Gamma\right]$$

$$\frac{\partial}{\partial w}\pi^T = \frac{\partial}{\partial w}[\pi_1 \cdots \pi_N]$$

$$= [\pi_1(1-\pi_1)x_1 \cdots \pi_N(1-\pi_N)x_N] = XR$$

let $R = \begin{bmatrix}\pi_1(1-\pi_1) & & 0\\ & \ddots & \\ 0 & & \pi_N(1-\pi_N)\end{bmatrix}$

$$\frac{\partial}{\partial w} w^T = \begin{bmatrix} \frac{\partial}{\partial w_1} \\ \vdots \\ \frac{\partial}{\partial w_d} \end{bmatrix} [w_1 \cdots w_d] = I$$

$$\Rightarrow \boxed{\nabla^2 E(w) = XRX^T + \Gamma}$$

d) Apply Newton-Raphson

$$w^{(new)} = w^{(old)} - \left[\nabla^2 E(w)\right]^{-1} \nabla E(w)$$

$$= w^{(old)} - \underbrace{(XRX^T + \Gamma)^{-1}}\left(X(\pi - y) + \Gamma w^{(old)}\right)$$

$(XRX^T+\Gamma)^{-1}(XRX^T+\Gamma)$

$$= (XRX^T + \Gamma)^{-1}\left((XRX^T + \cancel{\Gamma})\, w^{(old)} - X(\pi - y) + \cancel{\Gamma w^{(old)}}\right)$$

$$= (XRX^T + \Gamma)^{-1}\left(XRX^T w^{(old)} - X(\pi - y)\right)$$

RR^{-1}

$$\boxed{w^{(new)} = (XRX^T + \Gamma)^{-1} XR \underbrace{\left(X^T w^{(old)} - R^{-1}(\pi - y)\right)}_{z}}$$

IRLS