Taylor's Series:

$$f(x) = f(0) + f'(0)\cdot\frac{x}{1!} + f''(0)\cdot\frac{x^2}{2!} + f'''(0)\cdot\frac{x^3}{3!} + \dots$$

$$f(x) = f(a) + f'(a)\cdot\frac{(x-a)}{1!} + f''(a)\cdot\frac{(x-a)^2}{2!} + \dots$$

$$f(x) = \frac{\partial}{\partial x}\int_0^x f(x)\,dx$$ [Fundamental Theory of Calculus]

$$f(x) = \frac{\partial F_{area}}{\partial x}$$

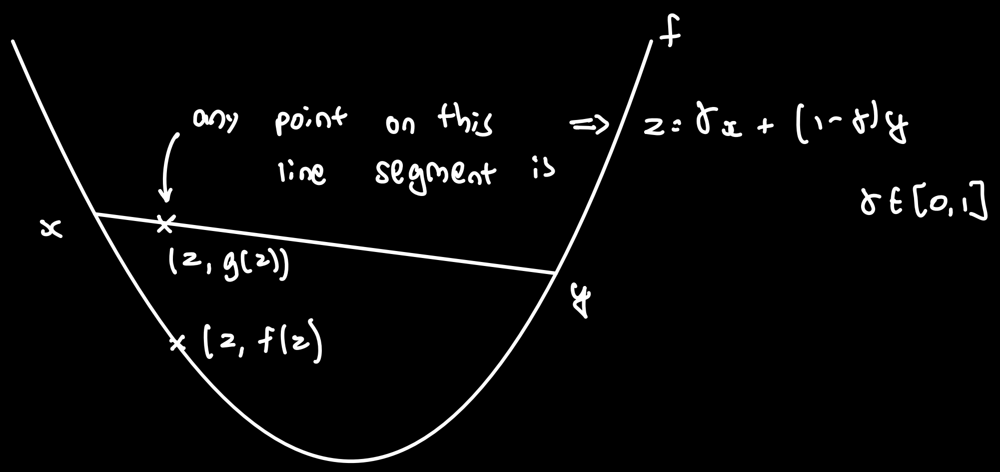
$$f(x) = f(0) + \int_0^x f'(t)\,dt$$

$$= f(0) + f'(0)\cdot x + \underbrace{\int_0^x (x-t)\cdot f''(t)\,dt}_{\text{remainder}}$$

$$f(x) = f(0) + f'(0)\,.\,(x) + f''(0)\,.\,(x^2/2) + \int_0^x (x-t)^2/_2\ f'''(0)\,.\,dt$$

Convex Functions:

$$f(\gamma x + (1-\gamma)y) \leq \gamma f(x) + (1-\gamma) f(y)$$



Whats $g(z)$?

Its a line segment so,

$$z = \gamma x + (1-\gamma) y$$

$$g(z) = \gamma g(x) + (1-\gamma) g(y)$$

$$= \gamma f(x) + (1-\gamma) f(y)$$

So, we are saying

f is convex if

$$f(z) \leq g(z)$$

Convex ⇒ Local minimum is Global minimum.

Assume x is local minimum

⇒ Any point in neighbourhood around x has larger value.

Assume y to be any point in the Domain.

A very close point to x can be

$(1-\delta)x + \delta y$ where δ is small positive number.

$f(x) \leq f((1-\delta)x + \delta y)$ Local Minimum — ①

By Convexity

$f((1-\delta)x + \delta y) \leq (1-\delta)f(x) + \delta f(y)$ — ②

Combine ①, ②

$f(x) \leq (1-\delta)f(x) + \delta f(y)$

$\delta f(x) \leq \delta f(y)$ $\delta \in [0,1]$

$f(x) \leq f(y) \rightarrow$ GLOBAL MINIMUM.

SMOOTHNESS:

"Self tuning". We want $\|\nabla f(x)\| \to 0$ as $x \to x^*$

i.e. $\|\nabla f(x) - \nabla f(y)\|_2 \leq \beta \|x - y\|_2$

"NO SHARP CHANGE IN GRADIENT".

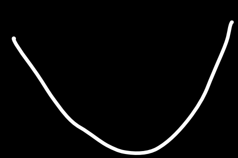
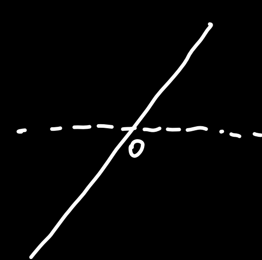
Slowly moves to x^*.

CLAIM:

If $f(x)$ is β-smooth

$\frac{\beta}{2}\|x\|_2^2 - f(x)$ is convex.

Note: f is convex if f' is monotonic increasing.

i.e. $f =$  $f' \equiv$ 

if f' is monotonic increasing

$y \geq x$ then either $f'(y) \geq f'(x)$ for all x, y

i.e. $\langle f'(y) - f'(x), y - x \rangle \geq 0$

So to prove $g(x) = \frac{\beta}{2}\|x\|_2^2 - f(x)$ is convex

we can prove $g'(x)$ is monotone

$$\langle g'(x) - g'(y), x-y \rangle \geq 0$$

$$\hookrightarrow \beta x - f'(x) - \beta y + f'(y)$$

$$\langle g'(x) - g'(y), x-y \rangle = \langle \beta(x-y) - (f'(x) - f'(y)), x-y \rangle$$

$$= \beta\|x-y\|_2^2 - \langle f'(x) - f'(y), x-y \rangle$$

Note: $\langle x,y \rangle \leq \|x\|_2 \|y\|$

$$\geq \beta\|x-y\|_2^2 - \|f'(x) - f'(y)\|^2 \|x-y\|$$

β - SMOOTH $\rightarrow$ $\|\nabla f(x) - \nabla f(y)\|_2 \leq \beta\|x-y\|_2$

$$\geq \beta\|x-y\|_2^2 - \beta\|x-y\|_2^2$$

$$\geq 0 \text{ //}$$

If $g(x)$ is convex then

$$g(y) \geq g(x) + \langle \nabla g(x), y-x \rangle$$

$$\frac{\beta}{2}\|y\|_2^2 - f(y) \geq \frac{\beta}{2}\|x\|_2^2 - f(x) + \langle \beta x - \nabla f(x), y-x \rangle$$

$$f(y) \leq f(x) + \frac{\beta}{2}\left(\|y\|_2^2 - \|x\|_2^2\right) - \langle \beta x - \nabla f(x), y-x \rangle$$

$$\Downarrow$$

$$\beta\langle x, y\rangle - \beta\|x\|_2^2$$
$$- \langle \nabla f(x), y-x \rangle$$

$$\leq f(x) + \frac{\beta}{2}\|y\|_2^2 - \frac{\beta}{2}\|x\|_2^2 + \beta\|x\|_2^2 - \beta\langle x, y\rangle$$
$$+ \langle \nabla f(x), y-x \rangle$$

$$\leq f(x) + \langle \nabla f(x), y-x \rangle + \frac{\beta}{2}\|y-x\|_2^2 \quad //$$

MONOTONICITY OF GD:

f is β - smooth and $\eta \leq 1/\beta$

$$f(x_{i+1}) \leq f(x_i) - \frac{\eta}{2} \|\nabla f(x_i)\|^2$$

UNIVARIATE:

$$f : \mathbb{R} \to \mathbb{R}$$

SMOOTHNESS UPPERBOUND:

$$\boxed{\forall a, b \qquad f(b) \leq f(a) + f'(a)(b-a) + \frac{\beta}{2}(b-a)^2}$$

$$f(x+h) = f(x) + f'(x) \cdot h + f''(x) \cdot \frac{h^2}{2} + \cdots$$

$$= f(x) + f'(x) \cdot h + \int_0^1 (f'(x+th) - f'(x)) \, th \, dt$$

$$f(b) = f(a) + f'(a) \cdot (b-a) + \int_0^1 (f'(a + t(b-a)) - f'(a)) \, t(b-a) \, dt$$

MONOTONICITY:

$$f(x_{i+1}) = f(x_i - \eta f'(x_i))$$

SMOOTHNESS UPPERBOUND

$$b = x_i - \eta f'(x_i)$$

$$a = x_i$$

$$\leq f(x_i) + f'(x_i)(-\eta f'(x_i)) + \frac{\beta}{2}(\eta^2)(f'(x_i))^2$$

$$= f(x_i) - \eta f'(x_i)^2 + \frac{\eta^2 \beta}{2}(f'(x_i))^2$$

$$= f(x_i) - \eta\left(1 - \frac{\eta\beta}{2}\right) f'(x_i)^2$$

$$\eta \leq \frac{1}{\beta}$$

$$= f(x_i) - \frac{\eta}{2} f'(x_i)^2$$

MULTIVARIATE:

$$f(y) \leq f(x) + \langle \nabla f(x), y-x \rangle + \frac{\beta}{2} \|y-x\|_2^2$$

$$f(x_{i+1}) = f(x_i - \eta \nabla f(x_i))$$

$$\leq f(x_i) + \langle \nabla f(x_i), -\eta \nabla f(x)_i \rangle + \frac{\beta}{2} \|(-\eta \nabla f(x_i))\|_2^2$$

$$\leq f(x_i) - \eta \|\nabla f(x_i)\|_2^2 + \frac{\eta^2 \beta}{2} \|\nabla f(x_i)\|_2^2$$

$$\leq f(x_i) - \eta \left(1 - \frac{\eta \beta}{2}\right) . \|\nabla f(x_i)\|_2^2$$

$$\leq f(x_i) - \frac{\eta}{2} \|\nabla f(x_i)\|_2$$

$$f(\omega_{k+1}) \leq f(\omega_k) - \frac{\eta}{2} \|\nabla f(\omega_k)\|^2 \qquad \eta = \frac{1}{\beta}$$

$$\leq f(\omega_k) - \frac{1}{2\beta} \|\nabla f(\omega_k)\|^2$$

$$\|\nabla f(w_k)\| \leq \varepsilon$$

$$\|\nabla f(\omega_k)\|^2 \leq 2\beta\,(f(\omega_k) - f(\omega_{k+1}))$$

UPTO ITERATION t NORM

$$\underbrace{\sum_{k=1}^{t} \|\nabla f(\omega_k)\|^2}_{\substack{\text{At least as}\\ \text{big as minimum}}} \leq 2\beta \sum_{k=1}^{t} (f(\omega_k) - f(\omega_{k+1}))$$

sum $\geq$ $t \times$ min

RHS

$$\sum_{k=1}^{t} [f(W_k) - f(\omega_{k+1})) = f(w_0) - f(\omega_1) + f(\omega_1) - f(\omega_2) \ldots - f(\omega_{t+1})$$

$$= f(w_0) - f(\omega_{t+1})$$

$\underset{=}{\text{LHS}}$

$$t \times \min_{j \in \{1, \ldots, t\}} \{ \| \nabla f(w_j) \|^2 \} \leq 2\beta (f(w_0) - f(w_{t+1}))$$

$$f(w^{t+1}) \geq f^*$$

$$\leq 2\beta (f(w_0) - f^*)$$

$$\min_{j \in \{1, \ldots, t\}} \{ \| \nabla f(w_j) \|^2 \} \leq \frac{2\beta}{t} (f(w_0) - f^*)$$

So $$\min_{j \in \{1, \ldots, t\}} \{ \| \nabla f(w_j) \|^2 \} \leq \varepsilon$$

if

$$\frac{2\beta}{t} (f(w_0) - f^*) \leq \varepsilon$$

$$t \geq \frac{2\beta (f(w_0 - f^*)}{\varepsilon}$$

$$f(w_1) \leq f(w_0) - \frac{\eta}{2} \|\nabla f(w_0)\|^2$$

$$\leq f(w_0) - \frac{1}{2\beta} \|\nabla f(w_0)\|^2$$

$$f(w_2) \leq f(w_1) - \frac{1}{2\beta} \|\nabla f(w_1)\|^2$$

$$\leq f(w_0) - \frac{1}{2\beta} \|\nabla f(w_1)\|^2 - \frac{1}{2\beta} \|\nabla f(w_0)\|^2$$

$$f(w_{t+1}) \leq f(w_0) - \frac{1}{2\beta} \Big(\|\nabla f(w_0)\|^2 + \frac{1}{2\beta} \|\nabla f(w_1)\|^2 + \cdots + \frac{1}{2\beta} \|\nabla f(w_t)\|^2 \Big)$$

$$\leq f(w_0) - \frac{1}{2\beta} \times t \times \underbrace{\text{smallest of } \|\nabla f(w_k)\|^2}$$

has to $\geq f^*$

$$f^* - f(w_0) \leq \frac{-t}{2\beta} \times \min \|\nabla f(w_k)\|^2$$

$$\frac{2\beta\,(f(w_0) - f^*)}{t} \geq \min \|\nabla f(w_k)\|^2$$

We want

$$\min \|\nabla f(w_k)\|^2 \leq \varepsilon$$

This is always true for

$$\frac{2\beta(f(w_0) - f^*)}{t} \leq \varepsilon$$

$$t \geq \frac{2\beta(f(w_0) - f^*)}{\varepsilon} \quad //$$

β - SMOOTH AND CONVEX:

$$\eta \leq 1/\beta$$

$$f(x_k) \leq f(x_*) + \frac{2\beta \|x_0 - x_*\|}{k}$$

↳ number of iterations.

SO ALL THE NEEDED EQUATIONS:

Convexity: $\forall x, y \quad f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$

β-Smooth: $f(x_{k+1}) \leq f(x_k) - \frac{1}{2\beta} \|\nabla f(x_k)\|_2^2$

GD: $x_{k+1} = x_k - \eta \nabla f(x_k)$

VECTOR: $\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2 \langle u, v \rangle$

$|\langle u, v \rangle| \leq \|u\| \cdot \|v\|$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2} \|y - x\|_2^2$$

$$f(x_*) \geq f(x_k) + \langle \nabla f(x_k), x_* - x_k \rangle + \frac{\alpha}{2} \|x_* - x_k\|^2$$

$$f(x_k) \leq f(x_*) - \langle \nabla f(x_k), x_* - x_k \rangle - \frac{\alpha}{2} \| x_* - x_k \|^2$$

$\beta \to$ SMOOTH:

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2\beta} \|\nabla f(x_k)\|^2$$

$$f(x_{k+1}) \leq f(x_*) - \langle \nabla f(x_k), x_* - x_k \rangle - \frac{\alpha}{2} \|x_* - x_k\|^2 - \frac{1}{2\beta} \|\nabla f(x_k)\|^2$$

$$\nabla f(x_k) = \beta(x_k - x_{k+1})$$

$$f(x_{k+1}) - f(x_*) \leq \frac{1}{2} 2\langle \beta (x_k - x_{k+1}), x_k - x_* \rangle - \frac{\beta}{2} \|x_k - x_{k+1}\|^2 - \frac{\alpha}{2} \|x_* - x_k\|^2]$$

$$\langle x_k - x_{k+1}, x_k - x_* \rangle$$

$$\leq \frac{\beta}{2}\|x_k - x_{k+1}\|^2 + \frac{\beta}{2}\|x_k - x_*\|^2$$

$$- \frac{\beta}{2}\|x_* - x_{k+1}\|^2 - \frac{\beta}{2}\|x_k - x_{k+1}\|^2$$

$$- \frac{\alpha}{2}\|x_* - x_k\|^2$$

$$\leq \frac{(\beta - \alpha)}{2}\|x_k - x_*\|^2 - \frac{\beta}{2}\|x_{k+1} - x_*\|^2$$

$$f(x_k) \geq f(x_*) + \langle \nabla f(x_*), x_k - x_* \rangle + \frac{\alpha}{2}\|x_k - x_*\|^2$$

$$f(x_k) - f(x_*) \geq \frac{\alpha}{2}\|x_k - x_*\|^2$$

$$f(x_{k+1}) - f(x_*) \geq \frac{\alpha}{2}\|x_{k+1} - x_*\|^2$$

$$\frac{\alpha}{2}\|x_{k+1} - x_*\|^2 \leq \frac{(\beta - \alpha)}{2}\|x_k - x_*\|^2 - \frac{\beta}{2}\|x_{k+1} - x_*\|^2$$

$$\frac{(\alpha+\beta)}{\beta}\|x_{k+1}-x_*\|^2 \le \left(1-\frac{\alpha}{\beta}\right)\|x_k - x_*\|^2$$

$$\left(1+\frac{\alpha}{\beta}\right)\|x_{k+1}-x_*\|^2 \le \left(1-\frac{\alpha}{\beta}\right)\|x_k - x_*\|^2$$

$$\|x_{k+1}-x_*\|^2 \le \frac{(\beta-\alpha)}{\alpha+\beta}\underbrace{\|x_k - x_*\|^2}$$

$$\le \frac{(\beta-\alpha)}{\alpha+\beta}\frac{(\beta-\alpha)}{\alpha+\beta}\|x_{k-1}-x_*\|^2$$

$$\le \left(\frac{\beta-\alpha}{\alpha+\beta}\right)^{k+1}\|x_0 - x_*\|^2$$

$$\|x_k - x_*\|^2 \le \|x_0 - x_*\|^2\left(\frac{\beta-\alpha}{\alpha+\beta}\right)^k$$

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2\beta} \|\nabla f(x_k)\|^2$$

$$f(x_k) \geq f(x_*) + \frac{\alpha}{2} \|x_k - x_*\|^2$$

$$-f(x_{k+1}) + f(x_k) \geq \frac{1}{2\beta} \|\nabla f(x_k)\|^2$$

$$\nabla f(x_k) = \beta(x_k - x_{k+1})$$

$$f(x_k) - \left(f(x_* + \frac{\alpha}{2} \|x_{k+1} - x_*\|^2\right) \geq \frac{\beta}{2} \|x_k - x_{k-1}\|^2$$

$$f(x_k) - f(x_*) - \frac{\alpha}{2} \| x_{k+1} - x_* \|^2 \geq \frac{\beta}{2} \|x_k - x_{k-1}\|^2$$

$$f(x_k) - f(x_*) \geq \frac{\beta}{2} \|x_k - x_{k-1}\|^2 + \frac{\alpha}{2} \|x_{k+1} - x_*\|^2$$

$$L(w) = \frac{1}{n} \sum_{i=1}^{n} \ell(y_i, \sigma(\langle w_i, x_i \rangle))$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(-y \log\Big(\underbrace{\frac{1}{1+e^{-\underbrace{\langle w_i x_i \rangle}_{z}}}}_{a}\Big) - (1-y)\log\Big(1 - \frac{1}{1+e^{-\langle w_i x_i \rangle}}\Big)\right)$$

$$\frac{\partial L_i}{\partial \vec{w}} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial z}{\partial w}$$

$$\frac{\partial L_i}{\partial a} = \frac{\partial}{\partial a}\Big(\quad (-y_i \log a - (1-y_i)\log(1-a)\Big)$$

$$= \frac{-y_i}{a} + \frac{(1-y_i)}{(1-a)}$$

$$\frac{\partial a}{\partial z} = \frac{e^{-z}}{(1+e^{-z})^2} = a(1-a)$$

$$\frac{\partial z}{\partial w} = x_i$$

$$\frac{\partial L_i}{\partial w} = \left(\frac{-y}{a} + \frac{(1-y_i)}{1-a}\right) a(1-a)\, x_i$$

$$= \left(-y_i \quad (1-a) \quad + \quad (1-y_i)\, a\right) x$$

$= (a - y_i) x_i$

$$\frac{\partial L}{\partial w} = \frac{1}{n} \sum_{i=1}^{n} (a - y_i) x_i$$

$a = \sigma(\langle w_i\, x_i \rangle)$

$- y_i \log a_i \;-\; (1 - y_i) \log(1 - a_i)$

$X \to n \times d$

$w \to d \times 1$ $\qquad$ $y = n \times 1$

$Xw = n \times 1$

$n \times 1$ $\qquad$ $n \times d$

$d \times n$ $\qquad$ $n \times 1$

$d \times 1$

$d \times n$ $\qquad$ $n \times 1$