$$\underbrace{x_1 \; x_2 \ldots \; x_n}_{\downarrow} \in D$$

$y_1 \; y_2 \cdots y_n$

GD:

$$x_{t+1} = x_t - \eta \nabla f(x_t)$$

$$|f(x_1) - f(x_2)| \leq L \|x_1 - x_2\|_2$$

$$\|\nabla f(x_1) - \nabla f(x_2)\|_2 \leq \beta \|x_1 - x_2\|_2$$

Smoothness Upper bound:

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

Monotonicity: ✓

GD to smooth ↓

$$f(x_{i+1}) \leq f(x_i) - \frac{\eta}{2} \underline{\underline{\|\nabla f(x_i)\|_2^2}}$$

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

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

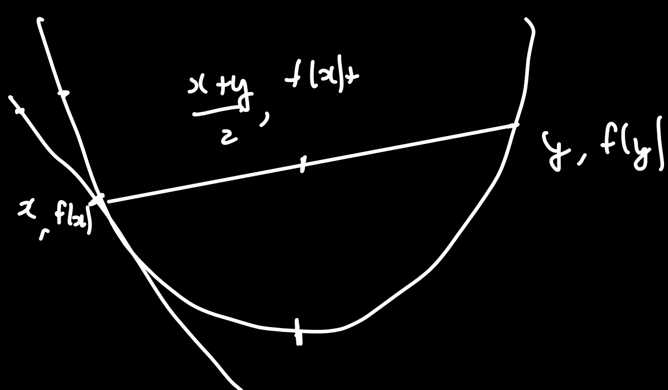
$$f(x_{i+1}) = f(\underbrace{x_i - \eta \nabla f(x_i)}_{b})) \qquad a = x_i$$

$$\leq f(x_i) + \langle \nabla f(x_i), -\eta \nabla f(x_i) \rangle$$

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

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

//



$$\Rightarrow f\left(\frac{x+y}{2}\right) \leq \frac{f(x) + f(y)}{2}$$

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$$

$$\boxed{f(x) + \langle \nabla f(x), y - x \rangle \le f(y)}$$

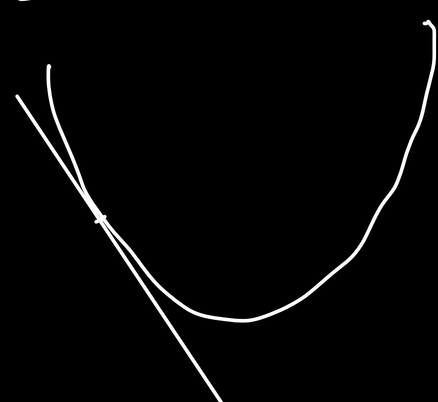
$$f(x_k) \le f(x_*) + \frac{2\beta \| x_0 - x_* \|}{k}$$

* $f(x) + \langle \nabla f(x), y - x \rangle \le f(y)$

* $f(x_{i+1}) \le f(x_i) - \frac{\eta}{2} \|\nabla f(x_i)\|_2^2$

$$f(x_{k+1}) \le f(x_k) - \frac{\eta}{2} \|\nabla f(x_k)\|_2^2$$

$$\underline{f(x_k)} + \langle \nabla f(x_k), x_* - x_k \rangle \le \underline{f(x_*)}$$



$$f(x_k) \ge f(x_{k+1}) + \frac{\eta}{2} \|\nabla f(x_k)\|_2^2$$

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

$$x_{t+1} = x_t - \eta \nabla f(x_t)$$

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

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

$$\uparrow \beta$$

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

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

$$\langle u, v \rangle = \frac{\|u\|^2 + \|v\|^2 - \|u - v\|^2}{2}$$

$$f(x_{k+1}) \leq f(x_*) + \underbrace{\frac{\beta}{2}\|x_k - x_{k+1}\|_2^2} + \frac{\beta}{2}\|x_* - x_k\|_2^2$$

$$- \frac{\beta}{2}\|x_* - x_{k+1}\|_2^2$$

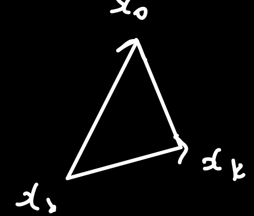
$$f(x_1) \leq f(x_*) + \frac{\beta}{2}\|x_* - x_0\|_2^2 - \underline{\frac{\beta}{2}\|x_* - x_1\|_2^2}$$

$$f(x_2) \leq f(x_*) + \underline{\frac{\beta}{2}\|x_* - x_1\|_2^2} - \underline{\frac{\beta}{2}\|x_* - x_2\|_2^2}$$

$$f(x_{k}) \leq f(x_*) + \underline{\frac{\beta}{2}\|x_* - x_{k-1}\|_2^2} - \frac{\beta}{2}\|x_* - x_k\|_2^2$$

$$\sum_{i=1}^{k} f(x_i) \leq k f(x_*) + \frac{\beta}{2}\left(\|x_* - x_0\|_2^2 - \|x_* - x_k\|_2^2\right)$$

$$f(x_k) \leq f(x_*) + \frac{\beta}{2k}\|x_0 - x_k\|_2^2$$



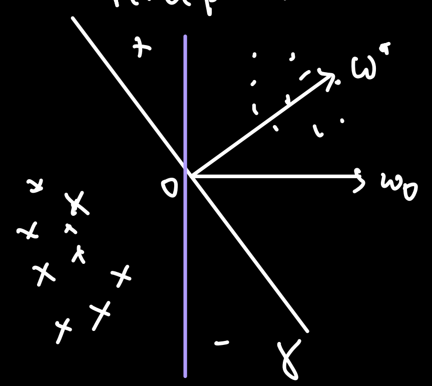
Online Learning

Mistake Bound

Regret Minimization

↓

perceptron



→ $|\langle x_i, w^*\rangle| \geq \gamma$

× $\boxed{w + y_i x_i}$

wrong

y_i — real answer

$y_{pred} = \text{sign}(\langle w_0, x_i\rangle)$

$\|w^*\| = \|w_0\| = 1$

$\|x_i\| = 1$

UPON MISTAKE:

$\|w_{i+1}\|^2 = \|w_i + y_i x_i\|^2$

$= \|w_i\|^2 + \|y_i x_i\|^2 + 2\langle w_i, y_i x_i\rangle$

$$y_i < w_i x_i >$$

$$= \|w_i\|^2 + 1 + \underbrace{\quad}_{\text{some -ve}} \quad - \quad +$$

$$\leq \|w_i\|^2 + 1$$

$$\|w_N\|^2 \leq \|w_0\|^2 + M_n$$

$$\|w_N\| \leq \sqrt{1 + M_n}$$

$$w_{i+1} = w_i + y_i x_i$$

$$y_i < x_i , w_* >$$

$$< w_{i+1} , w_* > = < w_i , w_* > + < y_i x_i , w_* >$$

$$< w_N , w_* > \geq < w_0 , w_* > + \gamma M_n$$

$$\|w_N\| \geq \gamma M_n - 1$$

$$\gamma M_n - 1 \leq \sqrt{1 + M_n}$$

$$M_n \leq 1/\gamma^2 + 1$$

REGRET MINIMIZATION:

$w_i = 1$ d experts

$$W_N \geq \left(\frac{1}{2}\right)^{L_*(T)}$$

On mistake $W_i \leq \frac{3}{4} W_{i-1}$

$$W_i \leq \left(\frac{3}{4}\right)^L W_0 \leq \left(\frac{3}{4}\right)^L d$$

$$\left(\frac{3}{4}\right)^L \cdot d \geq \left(\frac{1}{2}\right)^{L_*(T)}$$

$$d \cdot 2^{L_*(T)} \geq \left(\frac{4}{3}\right)^L$$

$$\log d + L_*(t) \geq L \log(4/3)$$

$$L \leq \frac{L_*(t)}{\log(4/3)} + \frac{\log d}{\log 4/3}$$

$$P(i) \propto w(t-1, i)$$

$$\Rightarrow \frac{w(t-1, i)}{\sum_j w(t-1, j)}$$

$$W(t, i) = (1-\varepsilon)\, w(t-1, i)$$

Expected loss at time t

$$\sum \Pr(\text{expert } i) \cdot \text{Loss of expert } i$$

$$L(t) = \sum_i \frac{e^{W(t-1, i)}}{\sum_j e^{W(t-1, j)}} \cdot L(t, i)$$

$$W(T) \geq (1-\varepsilon)^{A_*(t)} \cdot 1$$

$$W(T) \leq (1-\varepsilon)^{L(T)} \cdot d$$

$$1-\varepsilon \leq e^{-\varepsilon}$$

$$W(T) \leq e^{-\varepsilon A(T)} \cdot d$$

$$(1-\varepsilon)^{L_*(T)} \le e^{-\varepsilon A(T)} \cdot d$$

$$L_*(T) \ln(1-\varepsilon) \le -\varepsilon A(T) + \ln d$$

$$A(T) \le L_*(T) \frac{(-\ln(1-\varepsilon))}{\varepsilon} + \frac{\ln d}{\varepsilon}$$

$$l(t) = \sum_i \frac{w(t-1,i)}{\sum_j w(t-1,j)} \cdot l(t,i)$$

$$l(t) = \frac{1}{W(t-1)} \sum_i w(t-1,i) \cdot l(t,i)$$

$$W(t) = \sum_i w(t,i) = \sum_i (1-\varepsilon\, l(t,i))\, w(t-1,i)$$

$$= \sum_i w(t-1,i) - \varepsilon\, w(t-1,i)\, l(t,i)$$

$$= W(t-1) - \varepsilon\, l(t) \cdot W(t-1)$$

$$= W(t-1)(1-\varepsilon\, l(t))$$

$$\Pr \propto w(t-1, i)$$

$$L(t) = \frac{\sum_i L(t,i) \cdot w(t-1, i)}{w(t-1)}$$

$$w(t) = \sum_i w(t, i)$$

$$= \sum_i (1 - \varepsilon L(t,i))\, w(t-1, i)$$

$$= w(t-1) - \varepsilon L(t) \cdot w(t-1)$$

$$= w(t-1)[1 - \varepsilon L(t)]$$

$$\Pr[h(x^i) \neq y_i] \leq \gamma$$

$$T \geq \frac{2 \ln(1/\delta)}{(1/2 - \gamma)^2}$$

$$W(0) = 1$$

$$W(t) \leq W(t-1) \cdot (1-\varepsilon)^{L(t)}$$

$$W(T) \geq |BAD| \cdot (1-\varepsilon)^{T/2}$$

$$W(t) \leq W(t-1) \cdot (1-\varepsilon)^{1-\gamma}$$

$$L(t) \geq 1-\gamma$$

$$W(t) \leq W(t-1) \cdot (1-\varepsilon)^{L(t)}$$

$$\leq W(t-1) \cdot (1-\varepsilon)^{1-\gamma}$$

$$W(T) \leq d \cdot (1-\varepsilon)^{(1-\gamma)T}$$

$$W(T) \geq |BAD| \cdot (1-\varepsilon)^{T/2}$$

$$|BAD| \cdot (1-\varepsilon)^{T/2} \leq d(1-\varepsilon)^{(1-\gamma)T}$$

$$|BAD| \cdot e^{-\varepsilon T/2} \leq d\, e^{-\varepsilon(1-\gamma)T}$$

$$\frac{|BAD|}{d} \le e^{-\varepsilon T(1-\gamma-1/2)}$$

$$1/2-\gamma$$

S

$\langle x, v \rangle$ max

$\|X \cdot v\|^2$

$$X = U \Sigma V^T$$

$$X^T X = V \Sigma U^T U \Sigma V^T$$

$$= V \Sigma^2 V^T$$

$$XV = U\Sigma$$

$$X = \sigma_i u_i v_i^T + \dots$$

$$\sigma_i = \|X v_i\| \qquad X v_i = \sigma_i u_i$$

$$u_i = \frac{X v_i}{\sigma_i}$$

$$X^T X \cdot X^T X \cdot X^T X \cdot v_0$$

$$v = \frac{u}{\|u\|}$$

$$T = O\left(\frac{\log(d/\delta)}{\varepsilon}\right)$$

$$\|v_T - \mathrm{Proj}_{S_k}(v_T)\| \le \delta$$

$$k = 1$$

$$v_T - \mathrm{Proj}_{S_1}$$

$$\frac{\sigma_1}{2} > \sigma_2$$

$$X \to n \times d$$

$$\varepsilon = 1/2 \,/\!/ \qquad k = 1$$

$$\sigma_2 < (1-\varepsilon)\sigma_1$$

$$\|v_T - \mathrm{Proj}_{S_1}(v_T)\|^2$$

$$\downarrow$$

$$v_t = \frac{(X^T X)^T \cdot v_0}{\|(X^T X)^T \cdot v_0\|}$$

$$\sum_{i=1}^{d} \sigma_i^{2t} v_i v_i^T \cdot v_0$$

$$\sum_{i=1}^{d} \sigma_i^{2t} \langle v_i , v_0 \rangle \; v_i^T$$

$$\mathrm{Proj}_{v_1} \quad \sigma_1^{2t} \langle v_1 , v_0 \rangle \; v_1^T$$

$$\left\| \frac{\sum_{i=2}^{d} \sigma_i^{2t} \langle v_i , v_0 \rangle \; v_i^T}{\left\| \sum_{i=1}^{d} \sigma_i^{2t} \langle v_i , v_0 \rangle \; v_i^T \right\|} \right\|^2$$

$$\frac{(1-\varepsilon)^{2t} \sigma_1^{2t} \langle v_i , v_0 \rangle \; v_i^T \times (d-1)}{\sigma_1^{2t} \langle v_1 , v_0 \rangle \; v_1^T}$$

$$\leq \frac{\sigma_1^{4t} \; (1-\varepsilon)^{4t} \left\| \sum_{i=2}^{d} \langle v_i , v_0 \rangle \right\|^2}{\sigma_1^{4t} \langle v^1 , v_0 \rangle^2}$$

$$v_t$$

$$v_t = (X^T X)^t \cdot v_0$$

$$(X^T X)^t = \sum_{i=1}^{d} \sigma_i^{2t} \cdot v_i v_i^T$$

$$v_t = \sum_{i=1}^{d} \sigma_i^{2t} \langle v_i , v_0 \rangle v_i^T$$

$$\mathrm{Proj}_{v^1}(v_t) = \sigma_1^{2t} \langle v^1 , v_0 \rangle v^{1\,T}$$

$$\| v_t - \mathrm{Proj}_{v^1}(v_t) \|^2 = \| \sum_{i=2}^{d} \sigma_i^{2t} \langle v_i , v_0 \rangle \|^2$$

$$\leq \sigma_1^{4t} \cdot (1-\varepsilon)^{4t}$$

$$\| v_t \|^2 = \| \sum_{i=1}^{d} \sigma_i^{2t} \langle v_i , v_0 \rangle \|^2$$

$$\geq \sigma_1^{4t} | \langle v^1 , v_0 \rangle |^2$$

$$\leq \frac{(1-\varepsilon)^{2t}}{| \langle v^1 , v_0 \rangle |}$$

$$\frac{Pr[M(x)=a]}{Pr[M(x')=a]} \le \frac{1/2+\gamma}{\frac{1}{2}-\gamma} \qquad O(\gamma)$$

$$E(\gamma) = (1/2+\gamma)\, x_i + (1/2-\gamma)(1-x_i)$$

$$p(x) = \frac{1}{2b} e^{-|x|/b}$$

$f(x)+z=a$

$$\frac{Pr[M(x)=a]}{Pr[M(x')=a]} \qquad z = a - f(x)$$

$$= \frac{1/2b\; e^{-|a-f(x)|/b}}{1/2b\; e^{-|a-f(x')|/b}}$$

$$= e^{-\frac{1}{b}\left[|a-f(x)| - |a-f(x')|\right]}$$

$$= e^{-\frac{1}{b}\times s}$$

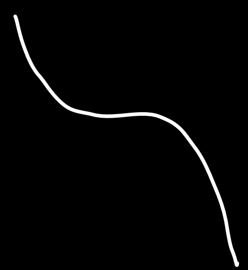
$$Pr[|z| > tb] \le e^{-t}$$

$$\frac{\Pr[u(M_\varepsilon(x)) = a]}{\Pr[u(M_\varepsilon(x')) = a]} = \frac{e^{\frac{\varepsilon}{\Delta u}[u(x,r)]}}{\sum_{r \in R} e^{\frac{\varepsilon}{\Delta u}(u(x,r))}} \times \frac{\sum_{r \in R} e^{\frac{\varepsilon}{\Delta u}(u(x',r))}}{e^{\frac{\varepsilon}{\Delta u}[u(x',r)]}}$$

$$u(x,r) - u(x',r) \leq \Delta u$$

$$e^{\varepsilon}$$

1. $f : \mathbb{R}^d \to \mathbb{R}$



$y = x_1^2 - x_2^2$

$\frac{\partial y}{\partial x_1} = 2x_1 \qquad \frac{\partial y}{\partial x_2} = -2x_2$

$(0,0)$

$\frac{\partial^2 y}{\partial x_1^2} = 2 \qquad \frac{\partial^2 y}{\partial x_2^2} = -2$

minima   maxima

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

$$f(w_{i+1}) \leq f(w_i) - \frac{1}{2\beta} \|\nabla f(w_i)\|_2^2$$

Let that point be w_k $\|\nabla f(w_k)\| \leq \varepsilon$

$$f(w_{k+1}) \leq f(w_k) - \frac{1}{2\beta} \|\nabla f(w_k)\|_2^2$$

$$\leq f(w_0) - \frac{1}{2\beta} \sum_{i=0}^{k} \|\nabla f(w_i)\|_2^2$$

$$\leq f(w_0) - \frac{1}{2\beta} \times k \, \|\nabla f(w_k)\|_2^2$$

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

$$k \leq \frac{[f(w_0) - f(w_*)]\,2\beta}{\varepsilon^2}$$

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

$$f(x_k) \leq f(x^*) + \frac{\beta}{2} \|x_k - x^*\|^2$$

$$f(x_0) \leq f(x^*) + \frac{\beta}{2} \|x_0 - x^*\|^2$$

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

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

$$\Rightarrow \quad f(x_{k+1}) \leq f(x_k) - \frac{\beta}{2} \|(x_k - x_{k+1})\|_2^2$$

$$\boxed{f(x_{k+1}) \geq f(x^*) + \frac{\alpha}{2} \|x_{k+1} - x^*\|_2^2}$$

$$f(x^*) + \frac{\alpha}{2} \|x_{k+1} - x^*\|_2^2 \leq f(x_k) - \frac{\beta}{2} \|x_k - x_{k+1}\|_2^2$$

$$f(x_k) \leq f(x^*) - \langle \nabla f(x_k), x^* - x_k \rangle$$

$$f(x_{k+1}) \leq f(x^*) + \beta \langle x_k - x_{k+1}, x_k - x^* \rangle - \frac{\beta}{2} \|x_k - x_{k+1}\|_2^2$$

$f(x^*) + \frac{\alpha}{2}\|x_{k+1} - x^*\|_2^2$

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

$$(\alpha+\beta)\|x_{k+1} - x^*\|_2^2 \leq \beta \|x_k - x^*\|^2$$

$$\|x_{k+1} - x^*\|_2^2 \leq \|x_k - x^*\|^2 \left(\frac{\beta}{\alpha+\beta}\right)$$

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

$$\leq \|x_0 - x^*\|^2 \left(\frac{\beta}{\alpha+\beta}\right)^k$$

$1 -$

0.9

$\alpha+\beta - \alpha$

$$\left(1 - \frac{\alpha}{\alpha+\beta}\right)^k$$

3. $f(x) \leq f(y)$

3. $Pr(t) \propto w(t-1, i)$

$$w(t,i) = (1-\varepsilon)^{L(t,i)} \, w(t-1,i)$$

$$w(T) \leq (1-\varepsilon)^{A(T)} n$$

$$w(T) \geq (1-\varepsilon)^{A_*(T)}$$

$$(1-\varepsilon)^{A_*(T)} \leq e^{-\varepsilon A(T)} \cdot n$$

$$A_*(T) \ln(1-\varepsilon) \leq -\varepsilon A(T) + \ln(n)$$

$$A(T) \leq \frac{\ln(n)}{\varepsilon} + \frac{(-\ln(1-\varepsilon))}{\varepsilon} A_*(T)$$

$$1+\varepsilon$$

$$\sum_i w(t,i) = \sum (1 - \varepsilon L(t,i)) \, w(t-1,i)$$

$$= w(t-1) - \varepsilon \sum L(t,i) \cdot w(t-1,i)$$

$$L(t,i) = \frac{\sum w(t-1,i) \, L(t-1,i)}{w(t-1)}$$

$$w(t-1) - \varepsilon L(t,i) \, w(t-1)$$

$$w(t) = w(t-1)(1-\varepsilon)^{L(t,i)}$$

$L = -y_i x_i$

$$w(t,i) = (1-\varepsilon)^{-y_i x_i} w(t-1,i)$$

$$w(T) \leq d.(1-\varepsilon)^{A(T)}$$

$$w(T) \geq 0$$

$$(1-\varepsilon)^{L(T)} \leq d.\, e^{-\varepsilon A(T)}$$

$$A(T) = \sum L(t)$$

$$0 \leq d(1-\varepsilon)^{A(T)}$$

$$0 \leq d\, e^{-\varepsilon A(T)}$$

$$\log d - \varepsilon A(t) \geq 0$$

$$A(T) \leq \log d / \varepsilon$$

1. $X = \sum \sigma_i u_i v_i^T$

$v = \sum \alpha_i v_i$

$X \cdot v = \sum \sigma_i u_i \alpha_i$

$\|X \cdot v\|^2 = \|\sum \alpha_i^2 \sigma_i^2\|$

$\leq \sigma_1^2 \|\sum \alpha_i^2\|$

$\|v\|^2$

v_1, v_2

w_1, w_2

var $\|X \cdot v_1\|^2 + \|X \cdot v_2\|^2$

$\geq$

$X^T X$ $\quad \underbrace{V \Sigma V^T}_{\sigma_1}$ $\quad$ 10, 9, 8

10 11 12

$2\sigma_1 I$

Let $B = OPT_u(x) - \Delta u(\ln|R| + t)/\varepsilon$

Let $r^* \Rightarrow u(x, r^*) = OPT_u(x)$

Some other r

$$\Pr[M_\varepsilon = r] = \frac{e^{\varepsilon \cdot u(x,r)/\Delta u}}{\sum e^{\varepsilon \cdot u(x,r)/\Delta u}}$$

$$\leq \frac{e^{\varepsilon \cdot u(x,r)/\Delta u}}{e^{\varepsilon \cdot u(x,r^*)/\Delta u}}$$

$$= e^{\varepsilon/\Delta u \cdot [\underbrace{u(x,r)}_{B} - \underbrace{u(x,r^*)}_{OPT_u(x)}]}$$

If $u(r) \leq B$

$$\leq e^{\varepsilon/\Delta u \times (-\Delta u)(\ln|R| + t)/\varepsilon}$$

$$= e^{-(\ln|R| + t)}$$

$$= \frac{e^{-t}}{|R|}$$

3. $(2k+1)$

$1 \quad 2 \quad N$ $\quad / \quad N-1$ $\qquad 1\ 2\ 3 \quad N-1 \quad N$

1

$k=0 \rightarrow N-1 \qquad N-2$

$k=1 \qquad N-3 \qquad 1\ 2$

$k=2 \qquad N-5$

$(N-1)-2k$

$$f(x_{i+1}) \leq f(x_i) - \frac{1}{2\beta}\|\nabla f(x_i)\|_2^2$$

$$f(x_k) + \langle \nabla f(x_k), x_k - x^* \rangle \leq f(x^*)$$

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

$$\leq f(x^*) - \langle \nabla f(x_k), x_k - x^* \rangle - \frac{1}{2\beta}\|\nabla f(x_k)\|_2^2$$

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

$$-\eta \nabla f(x_k)$$

$$x_{k+1} - x_k = -\eta\, G(x_k)$$

$$-\eta \langle \nabla f(x_k), \quad G(x_k) \rangle + \frac{\eta}{2} \| G(x_k) \|^2$$

$$E[f(x_{k+1})] \leq E[f(x_k)] - \eta \langle \nabla f(x_k), E(G(x_k)) \rangle + \frac{\eta}{2} E(\| G(x_k) \|^2)$$

$$+ \frac{\eta}{2} \| \nabla f(x_k) \|^2$$

$$+ \frac{\eta}{2} \sigma^2$$

$$E[f(x_{k+1})] \leq E[f(x_k)] - \frac{\eta}{2} E[\| \nabla f(x_k) \|_2^2] + \eta \frac{\sigma^2}{2}$$

$$E(f(x_k)) + E[\langle \nabla f(x_k), x_k - x^* \rangle] \leq E[f(x^*)]$$

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

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

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

$f(-x)$

$-f'(-x)$ $\quad$ $f''($

$f(-x)$

$y = -x$

$f(y)$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial x}$$

$$= -f'(y)$$

$f''(y)$

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

$$\leq f(x_*) + \langle \nabla f(x_k), x_k - x_* \rangle - \frac{1}{2\beta} \|\nabla f(x_k)\|_2^2$$

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

$$\sum f(x_{k+1}) \leq k f(x_*) + \frac{\beta}{2}\|x_0 - x_*\|^2 - \frac{\beta}{2}\|x_{k+1} - x_*\|^2$$

$$f(x_k) = 1.1 \; f(x_*)$$

$$0.1 \, f(x_*) \leq \frac{\|x_0 - x_*\|^2}{\beta \cdot k^2}$$

$$k^2 \leq \frac{\|x_0 - x_*\|^2}{\beta \cdot \varepsilon}$$

$$y^* = \operatorname{sign}(\langle w^*, x\rangle)$$

$$y = \operatorname{sign}(\langle w^*, x\rangle)$$

$$y^* \langle w^*, x\rangle \quad > 0$$

$$< 0$$

$$w_i = w_{i-1} + y_i x_i$$

$$\|w_i\|^2 = \|w_{i-1} + y_i x_i\|^2 = \|w_{i-1}\|^2 + \|y_i x_i\|^2$$

$$+ 2 \langle w_{i-1}, y_i x_i \rangle$$

$$y_i \langle w_{i-1}, x_i \rangle$$

$$\le 0$$

$$\le \|w_{i-1}\|^2 + 1$$

$$\|w_n\|^2 \le \|w_0\|^2 + M_n$$

$$\le 1 + M_n$$

$$\langle w_i, w^* \rangle = \langle w_{i-1}, w^* \rangle + y_i \langle x_i, w^* \rangle$$

$$\ge \langle w_{i-1}, w^* \rangle + \gamma \qquad \ge \gamma$$

$$\ge \langle w_0, w^* \rangle + M_n \gamma$$

$$\|w_n\| . \|w^*\| \ge \underbrace{\langle w_0, w^* \rangle}_{} + \gamma M_n$$

$$\ge \gamma M_n - 1$$

$$\gamma M_n - 1 \le \sqrt{1 + M_n}$$

$$Pr[i] = \frac{w(t,i)}{\sum w(t,i)}$$

$$L(t) = \sum Pr[i] \cdot L(t,i)$$

$$= \sum \frac{w(t-1,i)}{\sum w(t-1,i)} \cdot L(t,i)$$

$$= \frac{\sum w(t-1,i) \cdot L(t,i)}{w(t-1)}$$

$$w(t) = \sum_{i=1}^{d} (1 - \varepsilon L(t,i))\, w(t-1,i)$$

$$= w(t-1)\left[1 - \varepsilon L(t)\right]$$

$$w(T) \leq w(0)\,(1-\varepsilon)^{A(T)}$$

$$w(T) \leq d \cdot (1-\varepsilon)^{A(T)} \leq d \cdot e^{-\varepsilon A(T)}$$

$$w(T) \geq |BAD| \cdot (1-\varepsilon)^{T/2}$$

$$d \cdot e^{-\varepsilon T(1-\gamma)} \geq |BAD| \cdot (1-\varepsilon)^{T/2}$$

$$\frac{|BAD|}{d} \leq e^{-\varepsilon T(1-\gamma)} \cdot (1-\varepsilon)^{-T/2}$$

$$\left(\frac{1}{1-\varepsilon}\right)^{T/2}$$

$$e^{\varepsilon(1+\varepsilon)T/2}$$

$$v_t = \frac{\sum_{i=1}^{d} \sigma_i^{2t} \langle v^i, v_0 \rangle v^{iT}}{\sum_{i=1}^{d} \sigma_i^{2t} \langle v^i, v_0 \rangle}$$

$$\frac{\sigma_1^{4t} (1-\varepsilon)^{4t} \sum_{i=2}^{d} \langle v^i, v_0 \rangle}{\sigma_1^{4t} \langle v^1, v_0 \rangle^2}$$

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

$$f(x_k) \leq f(x_*) + \langle \underbrace{\nabla f(x_k)}_{\beta(x_k - x_{k+1})}, x_k - x_* \rangle$$

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

$$f(x_{k+1}) \leq f(x_*) + \frac{\beta}{2(k+1)} \|x_0 - x_*\|^2$$

$$P(x_1, x_2) = P(x_1) P(x_2)$$

$$P(x_1, x_2, x_3, x_4) = P(x_1, x_2, x_3 \mid x_4) P(x_4)$$

$$\Downarrow$$

$$1/24$$

$$= P(x_1, x_2 \mid x_3, x_4) P(x_3, x_4)$$

$$\underbrace{P(x_1 \mid x_2 x_3 x_4)}_{1/3} \underbrace{P(x_2)}_{1/2} \underbrace{P(x_3)}_{1/2} \underbrace{P(x_4)}_{1/2}$$

$$P(x_2) = 1/2$$

$$P(x_3) = 1/2$$

$$P(x_4) = 1/2$$

$p(x_1) = 1/2$

$p(x_1 | x_2) = 1/4$

$z = (x_1 + x_2 + x_3)^2$