Gradient Descent #2

SGD, On Data

$$(X,Y) = \{(x_1,y_1), \dots, (x_N,y_N)\}$$

$x_i \in \mathbb{R}^d$

$$\{M_\alpha \mid \alpha \in \mathbb{R}^m\}$$

$y_i \in \mathbb{R}$

$$\underset{\alpha \in \mathbb{R}^m}{\operatorname{Argmin}} \; \underbrace{L(\overset{\text{fixed}}{X,Y}, M_\alpha)}_{f(\alpha)}$$

$$\underset{\alpha \in \mathbb{R}^m}{\operatorname{Argmin}} \; f(\alpha)$$

$$(X,Y) = \{(x_1, y_1), \dots, (x_N, y_N)\}$$

$$x_i \in \mathbb{R} \quad (d=1) \quad y_i \in \mathbb{R}$$

poly. reg $\quad p = 2$

$$\alpha = (\alpha_0, \alpha_1, \alpha_2) \in \mathbb{R}^3$$

$$x_i \longrightarrow (1, x_i, x_i^2)$$

$$= \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2$$

$$M_\alpha(x_i) = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 \quad = \langle \alpha, (1, x_i, x_i^2) \rangle$$

$$\underbrace{SSE(X, Y, M_\alpha)}_{f(\alpha)} = \sum_{i=1}^{N} (M_\alpha(x_i) - y_i)^2$$

$$= (y_i - M_\alpha(x_i))^2$$

$$f(\alpha) = \frac{1}{N}\sum_{i=1}^{N} f_i(\alpha)$$

$$f_i(\alpha) = \left(M_\alpha(x_i) - y_i\right)^2$$

$$\alpha^* = \underset{\alpha \in \mathbb{R}^3}{\operatorname{argmin}}\, f(\alpha)$$

$\alpha^{(0)} = \alpha^{start}$

repeat

$\alpha^{(k+1)} = \alpha^{(k)} - \frac{\gamma}{N} \nabla f(\alpha^{(k)})$ $\quad \gamma'$

untile (- - -)

$$f(\alpha) = \sum_{i=1}^{N} f_i(\alpha)$$

$$f_i(\alpha) = (M_\alpha(x_i) - y_i)^2$$

$$\alpha = (\alpha_0, \alpha_1, \alpha_2)$$

$N = 1$ $\quad (x_1, y_1)$

$$f(\alpha) = (M_\alpha(x_1) - y_1)^2$$

$$\nabla f(\alpha) = \left(\frac{\partial}{\partial \alpha_0} f(\alpha), \frac{\partial}{\partial \alpha_1} f(\alpha), \frac{\partial}{\partial \alpha_2} f(\alpha) \right)$$

$$f_i(\alpha) = (M_\alpha(x_1) - y_1)^2 \qquad \frac{\partial}{\partial \alpha_2} f = 2(M_\alpha(x_1) - y_1)\, x_1^2$$

$$\frac{\partial}{\partial \alpha_0} (M_a(x_1) - y_1)^2 = 2(M_\alpha(x_1) - y_1) \frac{\partial (M_\alpha(x_1) - y_1)}{\partial \alpha_0}$$

$$= 2(M_a(x_1) - y_1) \frac{\partial(\alpha_0 + \alpha_1 x_1 + \alpha_1 x_1^2 - y_1)}{\partial \alpha_0}$$

$$= 2(M_a(x_1) - y_1) \cdot 1$$

$$\frac{\partial}{\partial \alpha_1} (M_a(x_1) - y_1)^2 = 2(M_a(x_1) - y_1) \frac{\partial(\alpha_0 + \alpha_1 x_1 + \alpha_1 x_1^2 - y)}{\partial \alpha_1}$$

$$= 2(M_a(x_1) - y_1)\, x_1$$

$N=1$

$$\nabla f(\alpha) = 2\left(M_\alpha(x_1) - y_1\right)\left(1, x_1, x_1^2\right)$$

$$\alpha^{new} = \alpha^{old} - \gamma\, 2\left(M_{\alpha^{old}}(x_1) - y_1\right)\left(1, x_1, x_1^2\right)$$

$N>1$

decomposable

$$f(\alpha) = \sum_{i=1}^{N} f_i(\alpha)$$

$$f_i(\alpha) = \left(M_\alpha(x_i) - y_i\right)^2$$

$$\nabla f(\alpha) = \nabla \sum_{i=1}^{N} f_i(\alpha)$$

$$= \sum_{i=1}^{N} \nabla f_i(\alpha)$$

α

$$M_\alpha(x_i) = \langle \alpha, (1, x_i, x_i^2) \rangle$$

$$\nabla f_i(\alpha) = 2\left(M_\alpha(x_i) - y_i\right)\left(1, x_i, x_i^2\right)$$

$$\nabla f(\alpha) = \sum_{i=1}^{N} 2\left(M_\alpha(x_i) - y_i\right)\left(1, x_i, x_i^2\right)$$

$p = 1$

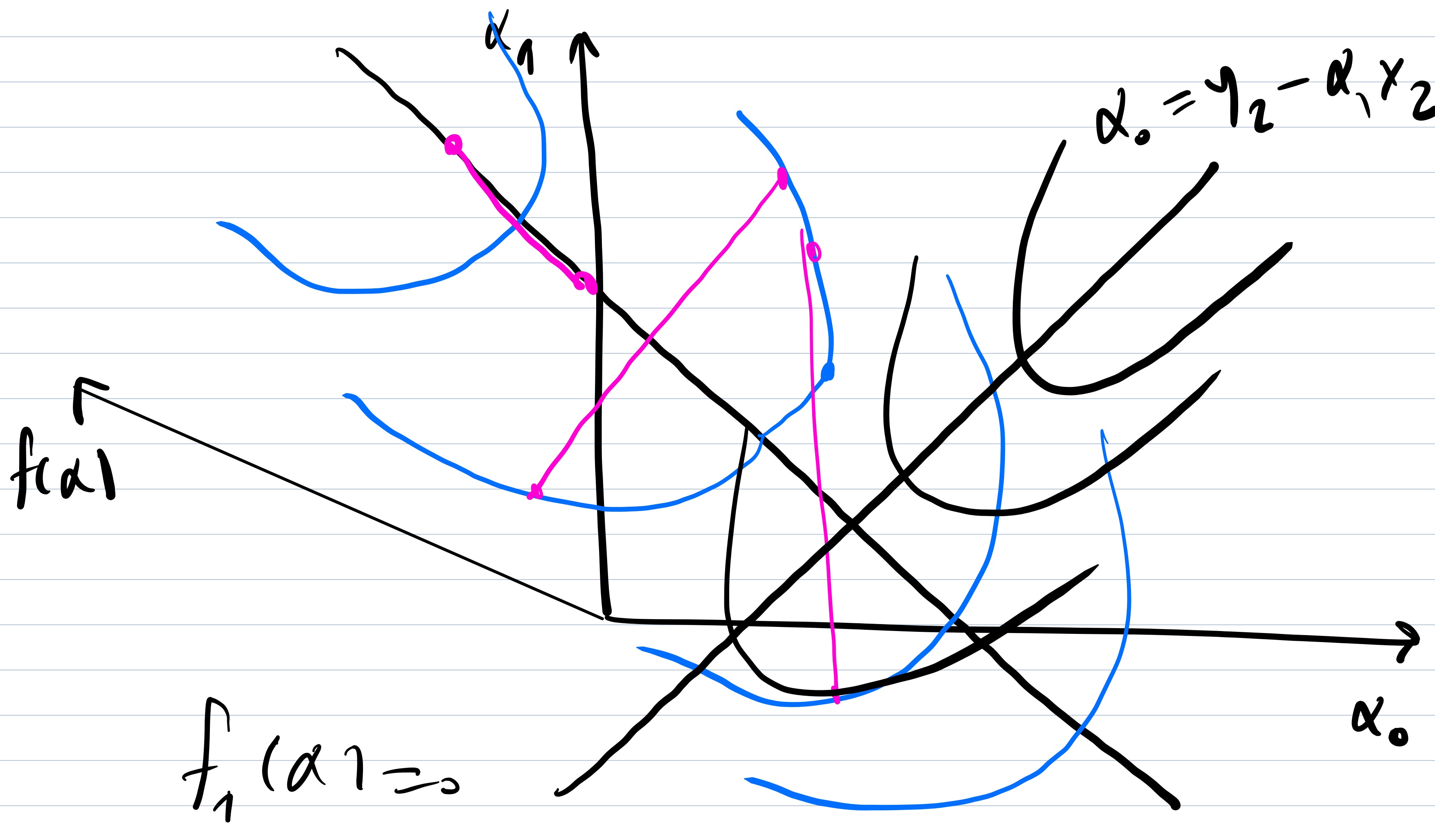
$$f(\alpha) = \sum_{i=1}^{N} f_i(\alpha) = \sum_{i=1}^{N} \left(M_\alpha(x_i) - y_i\right)^2$$

$$= \sum_{i=1}^{N} \underbrace{\left(\alpha_0 + \alpha_1 x_i - y_i\right)^2}_{= f_i(\alpha)}$$

if each f_i is convex, then $\sum f_i$ is convex

$$f_i(\alpha) = 0 \iff \alpha_0 + \alpha_1 x_i - y_i = 0$$

$$\iff \alpha_0 = y_i - \alpha x_i$$



$\alpha_0 = y_1 - \alpha_1 x_1$

$f_1(\alpha) = (\mu_\alpha(x_1) - y_1)^2$

$f(\alpha) = \sum_{i=1}^{N} f_i(\alpha)$ is strongly convex

if $N \geq 2$ (if at least two points are not non-general)

If $N \geq$ #parameters $\Rightarrow$ f is strongly convex.

$$\nabla f(\alpha) = \sum_{i=1}^{N} \nabla f_i(\alpha)$$

Incremental gradient descent

init. $\alpha^{(0)} = \alpha^{start}$

repeat

$$\alpha^{(k+1)} = \alpha^{(k)} - \gamma \nabla f_i(\alpha^{(k)})$$

$$i = i + 1 \pmod N$$

until $(\|\nabla f_i(\alpha)\| < T)$

$\hookrightarrow$ average of $\|\nabla f_i(\alpha)\| < T$ for some iterations.

Stochastic gradiant descent

initialize $\alpha^{(0)} = \alpha^{start}$

repeat

randomly $i \in \{1, 2 \ldots, N\}$

$$\alpha^{(k+1)} = \alpha^{(k)} - \gamma \nabla f_i(\alpha^{(k)})$$

until $(\|\nabla f_i(\alpha)\| < T)$

(*)

$$\nabla f(\alpha) = \frac{1}{N}\sum_{i=1}^{N} \nabla f_i(\alpha)$$

$$\mathbb{E}_{i\sim\{1,\dots,N\}}\left(\nabla f_i(\alpha)\right) = \nabla f(\alpha)$$