Computational Foundations for ML

10-607

Geoff Gordon

Notes and reminders

- Lab 0 due today (if submitting)

$$L(w) = \sum_{j=1}^{N} l_j(w)$$

ex. $\sum (w \cdot x_j - y_j)^2$

$$\frac{d}{dw} L(w) = \sum \frac{d}{dw} l_j(w)$$

\# iterations

for $t \leftarrow 1 \ldots T$:

\# batch

for $i \leftarrow 1 \ldots B$

\# training exs

$j_{ti} \leftarrow$ random $1 \ldots N$ $\Rightarrow$ w/o replacement

$g_{ti} \leftarrow \frac{d}{dw} l_{j_{ti}}(w_t)$

$g_t \leftarrow \frac{1}{B} \sum_{i=1}^{B} g_{ti}$

learning rate changes over iterations

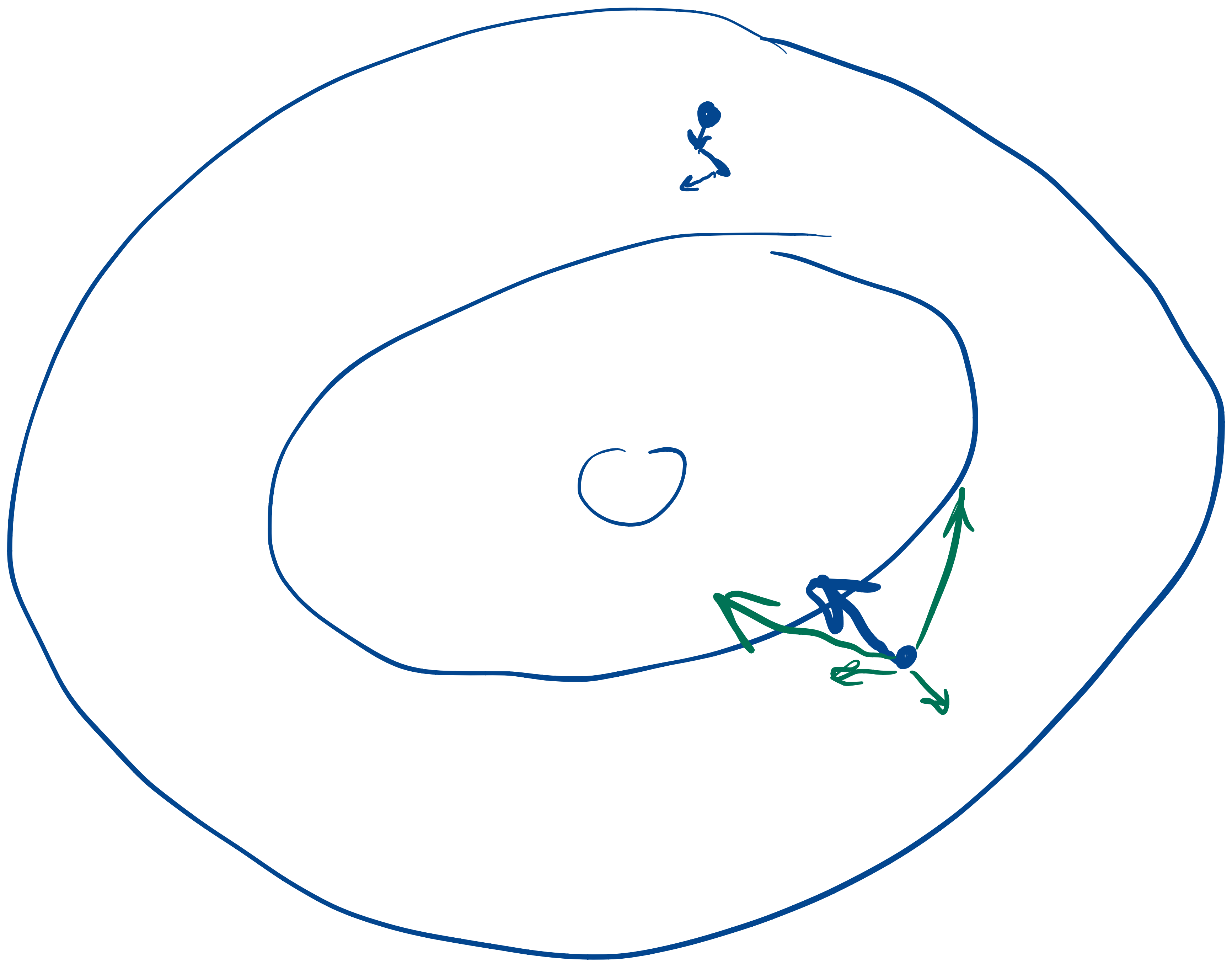
$w_{t+1} \leftarrow w_t - \eta_t g_t$

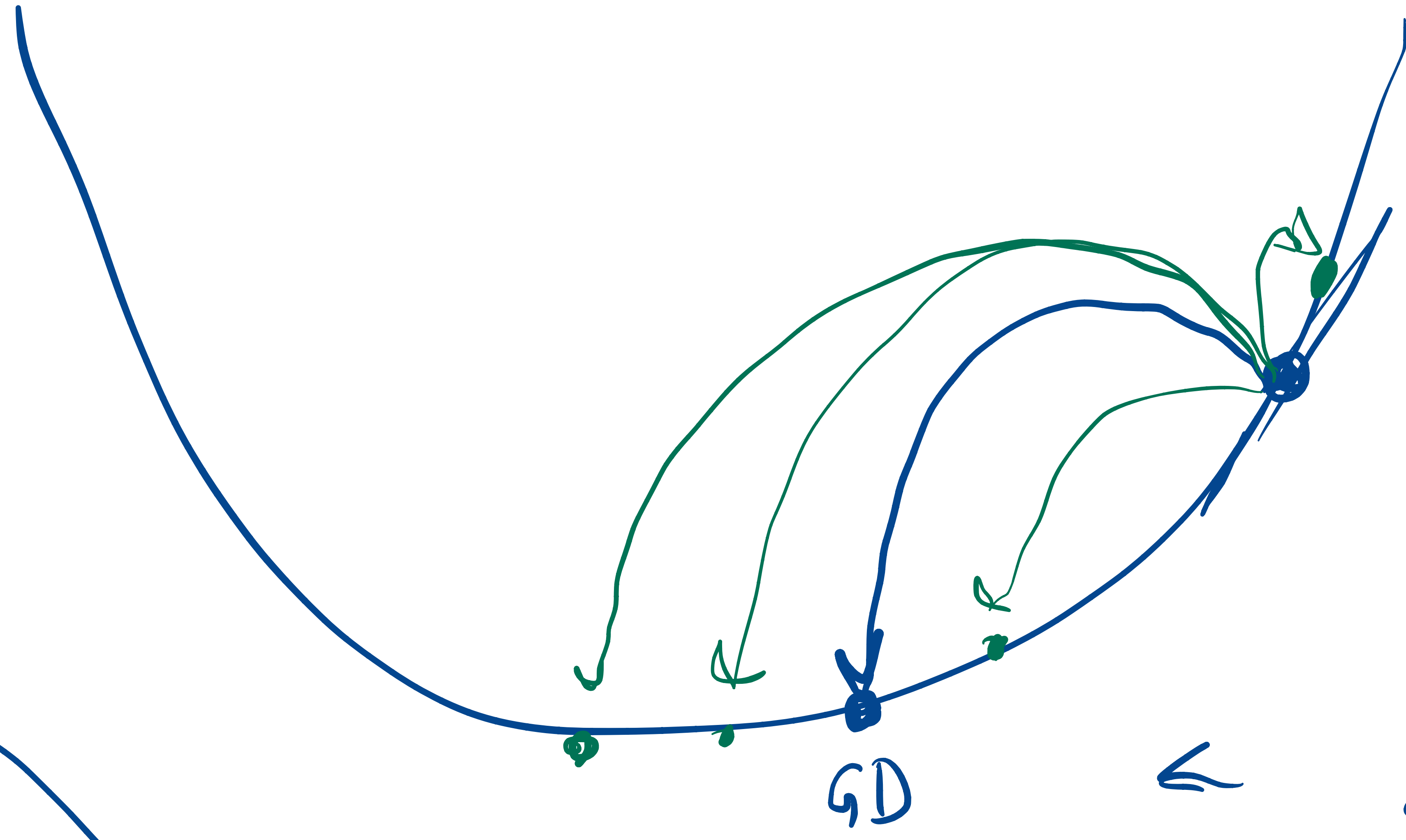
each iter uses B examples

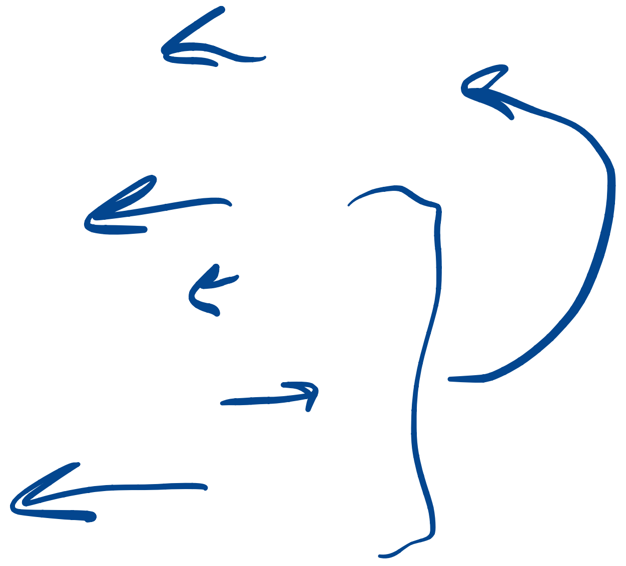
N total examples

$\left\lceil \frac{N}{B} \right\rceil$ iters in an epoch

$$E(L(w_{t+1})) \geq L(E(w_{t+1}))$$







true gradient

avg of example gradients

L∃∀N Coq Keymaera

$a, b, \text{happy}, \text{fuzzy}(\text{Spot}) \in \{T, F\}$

$a \to b \equiv \neg a \vee b$

$a \wedge b \quad a \vee b \quad \neg a \quad a \to b$

$(a \vee b) \wedge \neg c \to \text{happy}(\text{Spot})$

a	b	$a \to b$
F	F	T
F	T	T
T	F	F
T	T	T

modus ponens

$a \leftrightarrow b \equiv a \to b \wedge b \to a$

$$\frac{\begin{array}{l} a \\ a \to b \end{array}}{b} \quad \text{modus ponens}$$

$(a \wedge \text{fuzzy}(\text{Spot})) \vee z$

$$\frac{\begin{array}{l} \phi \\ \phi \to \psi \end{array}}{\psi} \quad \text{m.p.}$$

1. assume dog(Spot)
2. assume dog(Spot) $\to$ fuzzy(Spot)
3. conclude fuzzy(Spot) by m.p. from 1, 2

		premises	conclusion
$\wedge$:	intro:	ϕ ψ	$\phi \wedge \psi$
	elim:	$\phi \wedge \psi$	ϕ
		$\phi \wedge \psi$	ψ
$\vee$	intro:	ϕ	$\phi \vee \psi$ $\psi \vee \phi$
	elim	$\phi \vee \psi$ $\phi \to \chi$ $\psi \to \chi$	χ

premises	conclusion
—	T
F	ϕ

1.	assume $(a \wedge b) \wedge c$		
2.	c	$\wedge$-elim	from 1
3.	$a \wedge b$	"	" "
4.	a	"	" 3
5.	b	"	" "
6.	$b \wedge c$	$\wedge$-intro	5, 2
7.	$a \wedge (b \wedge c)$	"	4, 6

Part. exercise canceled for today

$$L(w) = \sum_{i=1}^{N} l_i(w) \qquad \text{e.g.} \quad L(w) = \sum_{i=1}^{N} (y_i - x_i \cdot w)^2$$

$$\frac{d}{dw} L(w) = \sum_i \frac{d}{dw} l_i(w) \qquad = -\sum_i 2(y_i - x_i \cdot w) x_i$$